Tax Evasion at the Top of the Income Distribution: Theory and Evidence *

John Guyton
Internal Revenue Service

Patrick Langetieg
Internal Revenue Service

Daniel Reck
University of Maryland

Max Risch
Carnegie Mellon University

Gabriel Zucman
Paris School of Economics and UC Berkeley

October 20, 2023

Abstract

We study tax evasion at the top of the U.S. income distribution using micro-data from random and operational audits and focused enforcement initiatives. Leveraging enforcement that revealed noncompliance ex post, we find that under the audit methods used during 2006–2013, individual random audit data failed to capture sophisticated evasion via offshore accounts and pass-through businesses. Consequently, estimates based solely on individual random audit data from this period under-state evasion by the highest-income Americans. We propose a theoretical explanation and construct new distributional estimates of noncompliance in the United States. Accounting for sophisticated evasion increases unreported income of the top 1% of the income distribution in 2006–2013 by 50% and increases the top 1% fiscal income share by about 1 percentage point.

*Corresponding Author: Daniel Reck, dreck@umd.edu. *We thank Gerald Auten, Pierre Bachas, Brian Galle, Bhanu Gupta, Anne Herlache, Tom Hertz, Xavier Jaravel, Drew Johns, Barry Johnson, Camille Landais, Katie Lim, Emily Lin, Larry May, Alicia Miller, Erik Ogilvie, Annette Portz, Ahmad Qadri, Peter Rose, Emmanuel Saez, Brenda Schafer, Clifford Scherwinski, Joel Slemrod, Matt Smith, Johannes Spinnewijn, David Splinter, Alex Turk, and Alex Yuskavage for helpful discussion, support, and comments on preliminary versions of this work. Francesco Armillei, Jeanne Bomare, and Baptiste Roux provided excellent research assistance. All remaining errors are our own. Financial support from the Washington Center for Equitable Growth, the Stone foundation, Arnold Ventures, and the Economic and Social Research Council is gratefully acknowledged.*

All findings, opinions, and errors are those of the authors alone and do not necessarily represent the opinions of the Internal Revenue Service or Treasury Department.

1 Introduction

How much do high-income individuals evade in taxes? And what are the main forms of tax noncompliance of the top of the income distribution? Because taxable income and tax liabilities are highly concentrated at the top of the income distribution, understanding noncompliance by high-income taxpayers is critical for the analysis of tax evasion, for tax enforcement, and for the conduct of tax policy.

A key difficulty in studying tax evasion by high-income individuals is the complexity of the forms of tax evasion available to them. Non-compliance by high earners often involves legal and financial intermediaries, such as shell companies and offshore banks, sometimes in countries with a great deal of secrecy. Moreover, a large fraction of the income of top earners derives from closely-held businesses. Misreporting in closely held businesses can be significant and yet difficult to detect, due to resource constraints within the tax authority, to the complexity of business structures with intertwined entities, and to the lack of third-party reporting. This complexity means that one single data source is unlikely to uncover all forms of noncompliance at the top of the income distribution.

In this paper, we attempt to quantify and overcome this limitation in the U.S. context by combining a wide array of sources of micro data, including (i) individual random audit data, (ii) the universe of operational audits conducted by the Internal Revenue Service (IRS), and (iii) focused enforcement activities (e.g., on offshore bank accounts and on specific evasive schemes used by private business owners). Drawing on this unique combination of data, we show that estimates based solely on individual random audits from 2006–2013 underestimate tax evasion at the top-end of the income distribution. Specifically, we find that under the audit procedures used during this period, individual random audit data do not capture most sophisticated evasion occurring via offshore intermediaries and pass-through businesses. We provide a theoretical explanation for these facts, and we propose a methodology to improve the estimation of the size and distribution of tax noncompliance in the United States.

The starting point of our analysis is the IRS individual random audit program, known as the National Research Program (NRP). Stratified random audits are commonly used to study and measure the extent of tax evasion.[1] Researchers use random audit data to test theories of tax evasion (Kleven et al., 2011), and tax authorities use them to estimate the extent of tax evasion – the *tax gap* – and select audits (IRS, 2019). Analyzing what individual audits detect in a representative sample of the full population yields important insights for policymakers and researchers, but when it comes to tax gap estimation especially, the main difficulty is that general purpose audits may not detect all forms of tax evasion. The NRP individual random audits from 2006–2013 appear

[1] We use the term tax evasion in this paper to refer to unintentional and intentional noncompliance with tax obligations, and do not attempt to distinguish between them.

to be well-designed to detect simple and common forms of tax evasion, such as unreported self-employment income, overstated deductions, and the over-claiming of tax credits. But, we argue, these audits did not detect more sophisticated types of evasion, at least during the period we study, because doing so would require additional information, resources, and/or specialized staff.

Our first contribution is to quantify the limits of NRP random audit data spanning tax years 2006 to 2013 when it comes to estimating top-end evasion in the United States. We find that in NRP individual random audit data from this period, detected evasion declines sharply at the top of the income distribution, with less than 1% of income found by auditors to be under-reported in the top 0.01%. Our analysis uncovers two key limitations of these audits which can account for this drop-off: tax evasion through foreign intermediaries (e.g., undeclared foreign bank accounts) and tax evasion via private businesses (e.g., under-reported revenues or overstated expenses in a partnership). We stress that these limitations apply to NRP individual random audits during our sample period and the audit procedures they entailed; they are not necessarily fundamental limitations of random audit data.

We find that offshore tax evasion went almost entirely undetected in our random audit data.[2] To establish this result, we analyze the sample of U.S. taxpayers who disclosed hidden offshore assets in the context of specific enforcement initiatives conducted in 2009–2012. The very top of the income distribution is dramatically over-represented in this sample. We find that offshore evasion was rarely detected in the audits, even when it was in fact occurring.

We also find that tax evasion occurring in pass-through businesses (S corporations and partnerships) is highly under-detected in our individual random audit data. Pass-throughs are private businesses whose income is taxed as income to their owners; pass-through ownership and income are highly concentrated (Cooper et al., 2016). During our sample period, examiners auditing pass-through owners during an individual random audit usually did not examine the degree to which pass-through businesses duly reported their income, especially for complex businesses. Thus, while the taxable income of taxpayers in the bottom 99% of the income distribution is comprehensively examined, up to half of the taxable income earned at the very top is not comprehensively examined under current individual random audit procedures. Analyzing micro-data from a small-scale random audit program for S corporations, we show that evasion in these businesses is substantial: more than 20% of net business income is under-reported, five times more than the total pass-through business income under-reporting detected in individual random audits. We also show that the adoption of sophisticated tax evasion schemes involving pass-through entities—micro-captive insurance and syndicated conservation easements—is prevalent among high-income taxpayers.

Our second contribution is to propose corrected estimates of under-reported income through

[2]Our data cover the period prior to the collection of third-party reported information on foreign bank accounts, which started in 2014. We analyze how our results can inform thinking about post-2014 evasion in Section 6.

the income distribution —and to investigate the consequences of this under-reporting for the measurement of inequality. We do so by starting from evasion estimated from individual random audit data and proposing a correction for sophisticated evasion that goes undetected in these audits. Although our corrected series feature only slightly more evasion on aggregate than in the standard IRS methodology, our proposed adjustments have large effects at the top of the income distribution. Our adjustments increase estimated unreported income by less than a factor of 1.2 on aggregate, but by a factor of 1.5 for the top 1%. After these adjustments, we estimate that under-reported income as a fraction of true income (i.e., income that should legally be reported on tax returns) rises from about 10% in the bottom 90% of the income distribution to 16% in the top 1%. Out of this 16%, 6 percentage points correspond to sophisticated evasion undetected in random audits. We also show that accounting for under-reported income increases the top 1% fiscal income share. In our preferred estimates, this share rises from 20.3% before audit to 21.4% on average over 2006–2013. The result that accounting for tax evasion increases inequality is robust to a range of robustness tests and sensitivity analysis (for instance, it is robust to assuming zero offshore tax evasion).

Our third contribution is to explain why general-purpose audits are not uniformly able to detect noncompliance across the income distribution. We consider a model in which taxpayers can adopt a technology that reduces the probability that their evasion is detected at some cost. We show that adoption of this technology is likely to be concentrated at the top of the income distribution for two reasons. First, high-income taxpayers have a greater demand for sophisticated evasion strategies that reduce the probability of detection, provided that, for large incomes, (i) the desired rate of evasion does not become trivial, and (ii) the cost of adopting becomes a trivial share of income. This is true even holding the probability of audit by income fixed. Second, audit rates are significantly higher at the top than at the bottom of the distribution, making evasion that is less likely to be detected and corrected on audit more attractive at the top. We can also re-interpret the model to think about situations where the outcome of an audit, if it occurs, is uncertain. With this interpretation, for the same reasons as before, we show that high-income people are more likely to adopt positions in the "gray area" between legal avoidance and evasion. Our model provides an explanation for why high-income individuals may evade more taxes than average despite facing a higher audit probability, changing our understanding of tax evasion by high-income persons relative to the canonical Allingham and Sandmo (1972) model.

Related Literature These findings provide a new picture of the distribution and composition of tax non-compliance, with implications for tax enforcement and for the effective progressivity of the US tax system. To optimize tax enforcement, it is important to know where in the income distribution missing revenues can be recovered and their sources. For a complete understanding of how the tax system shapes incentives, and how it treats different taxpayers, it is necessary to complement information on the distribution and composition of reported income with that on unreported income.

Our findings also have implications for the academic literature on tax evasion. Recent research suggests that taxpayers seldom evade taxes on third-party-reported income (Kleven et al., 2011; Carrillo et al., 2017; Slemrod et al., 2017; IRS, 2019), and that deterring evasion where enforcement is not supported by third-party information requires increasing the audit rate, the penalty rate, or tax morale (Luttmer and Singhal, 2014). This characterization works well for the middle and the bottom of the income distribution. But it misses the importance of the concealment of evasion (even from auditors) at the top and the adoption of aggressive interpretations of tax laws. From a government revenue perspective, the top of the income distribution is the sub-population where understanding tax evasion is the most important, due to the concentration of income and the progressivity of the individual income tax.

Due to a lack of data, there is relatively little research about noncompliance by high-income individuals. Recent research sheds light on one such form of noncompliance, offshore tax evasion (Johannesen et al., 2020; Alstadsaeter et al., 2019; Londoño-Vélez and Ávila-Mahecha, 2021). Relative to this body of work our main contribution is to cover a wider range of sophisticated noncompliance, including most importantly noncompliance through private businesses, a key source of income at the top of the income distribution. This is particularly important in the US, where a high and rising fraction of taxable individual income at the top of the distribution flows from private pass-through businesses (Cooper et al., 2016; Smith et al., 2019a). As much as 50% of taxable income for the top 0.01% highest earners derives from such pass-throughs, a distinguishing feature of the US tax system. Understanding noncompliance involving pass-throughs is therefore essential to paint an accurate picture of the size and distribution of US individual income tax noncompliance. Our results suggest that the growth of pass-through structures since the 1980s coincided with the development of new strategies for sophisticated tax evasion, strategies facilitated by pass-through ownership structures. We find that circa 2007, evasion via pass-through businesses made a contribution to the tax gap that was perhaps twice as large as the contribution of offshore evasion. Since 2007, offshore evasion has been subjected to an ambitious international crackdown (De Simone et al., 2020; Casi et al., 2020), while, as we discuss further below, pass-through businesses have grown in importance and audits of pass-through businesses have fallen dramatically. Compared to earlier work, our results suggest that sophisticated tax evasion is broader and more systematic than the concealment of wealth in offshore accounts specifically, and that tax policy shapes the types of sophisticated evasion strategies that high-income, high-wealth persons adopt.

The rest of this paper is organized as follows. Section 2 studies the distribution of noncompliance detected in individual random audit data. Sections 3 and 4 provide direct evidence that offshore and pass-through evasion are (i) highly concentrated at the top of the income distribution, (ii) rarely uncovered in individual random audits, and (iii) quantitatively important for the measurement of income at the top. Section 5 considers other adjustments for undetected evasion.

In Section 6 we present our estimates of the distribution of noncompliance and investigate their implications for the measurement of inequality. Section 7 presents our theoretical analysis, and Section 8 concludes. The Appendix contains additional empirical and theoretical analysis.

2 The Distribution of Evasion Detected in Random Audits

2.1 Background on IRS Random Audits

The National Research Program random audits are the main data source used to study the extent and composition of individual tax evasion in the United States (see, e.g., Andreoni et al., 1998; Johns and Slemrod, 2010; IRS, 2016a, 2019; DeBacker et al., 2020).[3] NRP auditors assess compliance across the entire individual tax return—the Form 1040—based on information from the schedules of this Form, third-party information reports, the taxpayer's own records, and measures of risk comparing all this information to information on the broader filing population.[4] The procedures followed in the 2006–2013 random audits were standard audit procedures for audits of individual taxpayers conducted by the Small Business and Self-Employed operating division of the IRS.

Our analyses of the NRP pool data from the random audits conducted for tax years 2006–2013. The NRP uses a stratified random sample which over-samples top earners. Our pooled sample includes 105,167 audited taxpayers, of which 12,003 are in the top 1% of the reported income distribution. We use the NRP weights to compute statistics that are representative of the full population of individual income tax filers. Our sample is large enough to obtain precise estimates for groups as small as the top 0.01% (although splitting this top group by other characteristics tends to leave us with too little statistical power for informative analysis). Our main statistic of interest is the *rate of income under-reporting*—the amount of income under-reported expressed as a fraction of true income.[5] In the main text we focus on point estimates; confidence intervals for the rate of income mis-reporting are reported in the Online Appendix (Figure A1). As we are interested in assessing how noncompliance affects measured inequality, we use the same definition of income as Piketty and Saez (2003), market income defined as total income (Line 22 of the 2012 Form 1040) less Social Security benefits (Line 20b), unemployment compensation (Line 19), alimony (Line 11), and taxable refunds (Line 10). As we show below, other income definitions yield similar results.

It has long been acknowledged that in the context of an audit, some noncompliance may go

[3]Further background on the NRP is in the Internal Revenue Manuals here: `https://www.irs.gov/irm/part4/irm_04-022-001`.

[4]We use the terms "NRP audits" and "NRP auditors" to refer to audits conducted as part of the National Research Program. Earlier IRS random audit studies under the Taxpayer Compliance Measurement Program (TCMP) consisted of line-by-line examinations of the individual tax return. The NRP aims to provide a similarly comprehensive measure of compliance at a reduced administrative cost and burden on the taxpayer. See Brown and Mazur (2003) for more on the TCMP and how the NRP uses revised procedures to achieve similar objectives.

[5]Tax Gap studies (IRS, 2016a, 2019; Johns and Slemrod, 2010) often estimate a similar quantity called the Net Misreporting Percentage, income under-reporting divided by the total of the absolute value of true income, which can differ from what we estimate for components of income that can be negative. We use a different term here partly because we never use absolute values of negative components of income.

undetected. The IRS uses a methodology, known as Detection-Controlled Estimation (DCE), to estimate undetected noncompliance (Feinstein, 1991; Erard and Feinstein, 2011). Official estimates of the individual income tax gap for the period we study (e.g., IRS, 2019) use DCE methodology, as does prior work on the distribution of noncompliance (Johns and Slemrod, 2010). In this section, we describe what is detected in the course of individual random audits, without any correction for undetected noncompliance.[6] We incorporate the additional under-reporting identified by DCE into our analysis in Section 5; before this Section, all of our analysis excludes any DCE adjustment.

2.2 The Distribution of Detected Evasion

We start the analysis by showing the rate of income under-reporting across the income distribution. Unless otherwise noted, all our analyses in this paper rank taxpayers by their estimated true income (with different measures of "true income" depending on the method used to estimate unreported income).[7] On aggregate, 4.0% of true income is found under-reported in NRP random audits.[8] Figure 1a shows that the rate of income under-reporting hovers around 4% to 5% through most of the income distribution and falls sharply within the top 0.1% to less than 1% in the top 0.01%.[9][10] As shown by Appendix Figure A1, this profile of evasion is relatively precisely estimated.

Figure 1a also decomposes unreported income by type of income. By far the largest type of unreported income is Sole Proprietor income, reported on Schedule C of the income tax return. Under-reporting in this category comprises about 50% of all detected evasion. The next-largest category involves corrections to Form 1040 Line 21 income ("Other income"), which mostly reflect disallowed business losses carried over from previous years. Adjustments to line 21 are relatively uncommon but they can be large when they occur. The taxpayers with the largest adjustments in this category typically have negative reported income due to past business losses, but once those

[6]DeBacker et al. (2020) perform a similar analysis and obtain similar results. However, because of our subsequent results on offshore and business evasion, our interpretation of the patterns seen in the NRP is different. We argue that the low detected evasion at the top is a consequence of the fact that sophisticated evasion is less likely to be detected, not that high-income taxpayers are much more compliant.

[7]Ranking taxpayers by reported income would lead to downward-biased estimates of the rate of income under-reporting at the top, because the act of under-reporting income moves taxpayers down the distribution of reported income. Figure A3 and Table A2 illustrate how re-ranking from reported to corrected income shapes these estimates.

[8]This estimate is smaller than the official IRS (2019) estimates of noncompliance for this period, in which 11% of aggregate true income is unreported, because the official estimates include undetected noncompliance estimated by DCE methods.

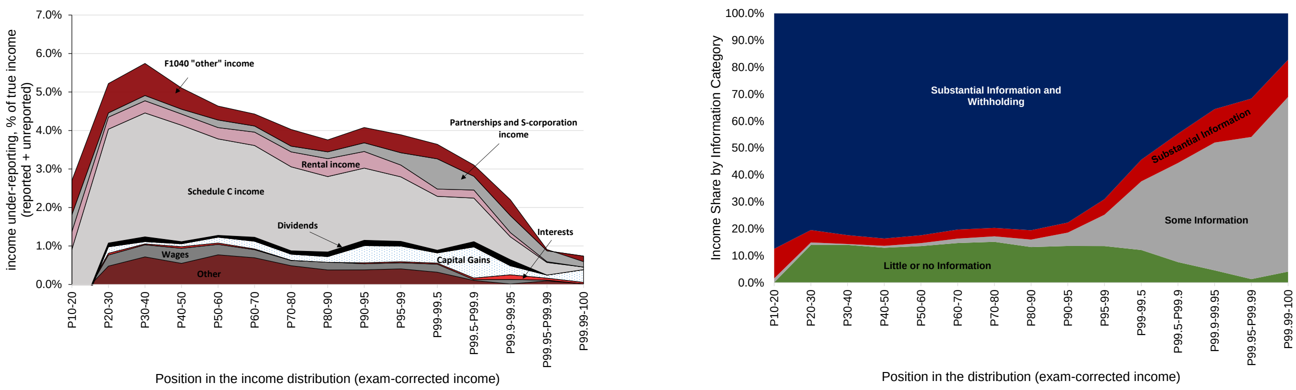
[9]Unless otherwise noted, our estimates include all taxpayers, including those found over-reporting income and those with no change in income upon audit. Taxpayers who under-reported income under-reported 4.5% of aggregate true income, while taxpayers who over-reported income over-reported the equivalent of 0.5% of aggregate true income. The majority of over-reported income is in the bottom half of the exam-corrected income distribution; the implications of over-reporting for aggregate tax liabilities and for noncompliance at the top are negligible.

[10]We omit the P0-P10 group in Figure 1a because aggregate exam-corrected income in this group is negative, which complicates the interpretation of the rate of under-reporting. This excludes 1.4% of total net under-reporting from depiction in the Figure. Although about 21% of total net under-reporting is found for individuals in the bottom bin of the *reported* income distribution, we show in Appendix Table A2 that the vast majority of this under-reporting is done by taxpayers who rank further up in the *exam-corrected* income distribution. This fact is also apparent in Figure A3.

losses are disallowed, their exam-corrected income falls in the top 1% of the income distribution. Additional information on the composition of under-reported income in the full population and in the top 1% is reported in Appendix Table A1.

FIGURE 1: DETECTED UNDER-REPORTING AND THE COMPOSITION OF INCOME ACCORDING TO RANDOM AUDIT DATA

(a) Detected Under-Reporting (b) Composition of Income



Notes: Figure 1a shows unreported income detected in 2006–2013 NRP individual random audit data, without any correction for undetected evasion, expressed as a fraction of exam-corrected income. Tax units are ranked by their exam-corrected market income (defined as total income reported on form 1040 minus Social Security benefits, unemployment insurance benefits, alimony, and state refunds). Detected unreported income decreases sharply within the top 1% of the income distribution. Misreporting of Schedule C income comprises the bulk of detected evasion. Little evasion is detected for partnership and S corporation business income and financial capital income, important sources of income at the top. Figure 1b plots the share of exam-corrected market income in each of the four categories used by IRS (2016a) to describe the comprehensiveness of third-party information and withholding. "Little or no information" includes non-farm proprietor income, line 21 other income, rents and royalties, farm income, and form 4797 income. "Substantial information" includes interest and dividends. "Some information" includes partnership and S corporation business income and capital gains. "Substantial information and withholding" includes just wage and salary income. Chart 1 of IRS (2016a) reports estimated under-reporting rates of 63% for "little or no information," 19% for "some information," 7% for "substantial information," and 1% for "substantial information and withholding." We observe in the figure a shift in the composition of income from the bottom 99% to the top of the top 1%, away from income with substantial information reporting and toward income with only limited information reporting (e.g. Schedule K-1 reporting for partnership sand S corporations, 1099-B reporting for some but not all capital gains).

Four remarks are in order. First, our results are robust to using other definitions of income. Appendix Figure A2 shows that using adjusted gross income rather than market income gives nearly identical results for all but the very bottom bin. Second, NRP random audits detect little evasion on wages, interest, dividends, and capital gains. A natural explanation for this fact is that these forms of income are subject to substantial third-party reporting (IRS, 2016a; Kleven et al., 2011).[11] However, interest, dividends, and capital gains accruing to offshore accounts only started being subject to information reporting with the implementation of the Foreign Account Tax Compliance Act in 2014, after our period of study. Thus, the low evasion rates on financial capital income recorded in the 2006–2013 NRP may, in part, be due to the fact that some evasion on offshore capital income

[11]Capital gains were subject to some information reporting throughout our period of study; a reform added the requirement that brokers report not just sale price but also cost basis starting in tax year 2011.

went undetected.

Third, there is an asymmetry in the detected rates of evasion across different types of business income. For sole proprietor income, which is supported by relatively little third-party information (Slemrod et al., 2017), the under-reporting rate is 37% overall and 19% for the top 1%. For partnership and S corporation business income, also supported by relatively little third-party information, detected noncompliance is much lower: 5% overall and 2% for the top 1%.[12] As discussed below, sole proprietor income is subject to extensive examination in the context of these audits, while the examination of S corporation and partnership business income faces practical limitations.

Fourth, detected evasion among those with very high incomes is extremely low. In the top 0.01% by exam-corrected income, just 0.6% of true income is under-reported. This is a direct consequence of the fact that the NRP uncovers little noncompliance on interest, dividends, pass-through business income, and capital gains—the key sources of income in the top 0.01%. This observation is especially surprising given the lack of comprehensive information reporting on key sources of income at the top. Figure 1b illustrates how the extent of information reporting changes across the income distribution. In the bottom 99% of the distribution, about 90% of market income is wage and pension income, which is subject to substantial information reporting and withholding. Within the top 1%, there is a dramatic shift away from wages and pensions, towards private business income and capital gains, for which there is only limited information reporting.[13] Prior literature suggests a strong empirical relationship between the presence of third-party information and rates of evasion (Kleven et al., 2011; IRS, 2019). IRS (2019) estimates that the rate of under-reporting on income subject to substantial information reporting and withholding is just 1%, while the rate of under-reporting on income subject to "some information" reporting is 17%. As such, we might expect the shift in the composition of income we observe in Figure 1b to be accompanied by an increase in rates of under-reporting, but we observe the opposite in Figure 1a.

2.3 Direct Evidence for Undetected Evasion at the Top

We now provide simple and direct evidence that our random audit data miss significant top-end evasion. We compare the amount of evasion estimated at the top in random audit data with the amount found in operational audits. Operational audits include all audits other than NRP random audits: correspondence audits, conventional in-person audits, sophisticated audits of high-income/high-wealth individuals, and a variety of other specialized audit programs. The IRS prioritizes audits of taxpayers who, based on a variety of factors, it expects to be non-compliant, so

[12] Schedule K-1 provides some information on pass-through business income, but whether this is truly "third-party" information depends on whether a given individual owner can control what is being reported on the K-1. Moreover, the K-1 amounts could already be understating income due to under-reporting on the business's tax return. We discuss both of these issues in Section 4.

[13] In the top 0.01%, about two-thirds of income derive from private businesses, capital gains, and "other income." An additional 15% derives from interest and dividends, some of which (e.g., investment income accruing to offshore accounts) were subject to limited information reporting during our sample period

operational audits are not typically conducted at random. On average between 2007-2013, 17% of tax units in the top 0.01% were audited per year.[14]

Because only a fraction of the population is subject to an operational audit in a given year, evasion uncovered in operational audit data should be much lower than the population-weighted NRP estimates of evasion. However, as shown by Figure 2, this is not the case at the top. The population estimate of evasion for the top 0.01% from the NRP falls substantially below the amount assessed in operational audits alone. On average over fiscal years 2007-2013, in operational audits $967 million (in 2012 dollars) per year were assessed in additional taxes owed by the top 0.01% (by reported income). In the extreme scenario where all non-audited taxpayers evaded zero, this implies that the top 0.01% by reported income evaded at least 1.2% of taxes owed. This lower bound for the amount of evasion in the top 0.01% is already far larger than the NRP point estimate, which is less than 0.05%.[15] In the next two bins—P99.9 to P99.95 and P99.95 to P99.99—the lower bound based on taxes assessed in operational audits is comparable to the full population NRP estimate, despite the fact that less than 10% of taxpayers in these groups were subject to an operational audit. These results provide direct evidence that NRP audits from our sample period must miss some noncompliance at the top.[16]

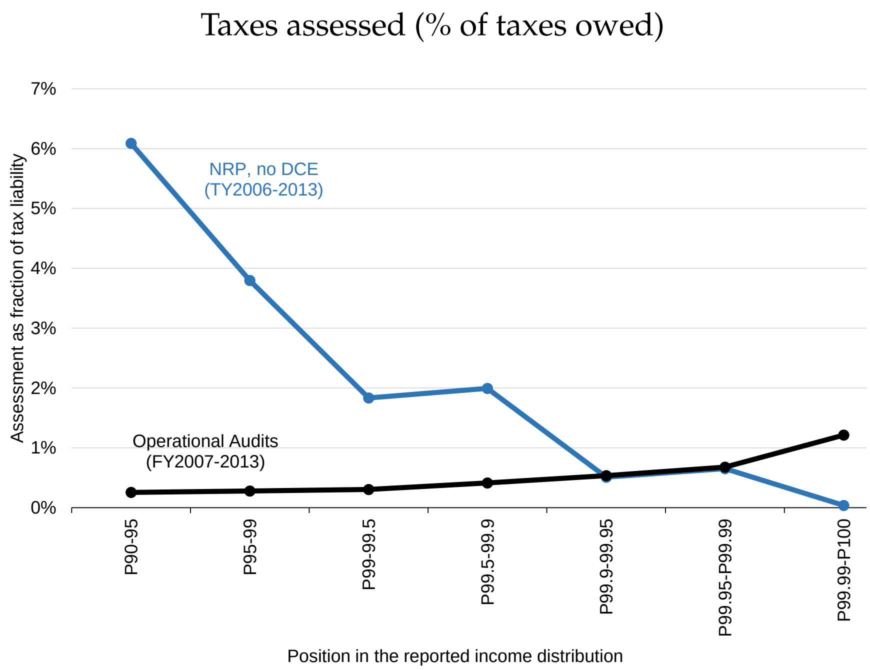
Random audit data from the period we study face two main limitations. First, the available tools, procedures, and resources placed limits on the extent to which some types of evasion are detected, especially at the top. Until recently, there was little evidence available on the nature and magnitude of these sophisticated types of evasion, making it hard to quantify them. We examine new evidence on this issue in Sections 3 (offshore evasion) and 4 (pass-through businesses). Second, even for a given set of tools and procedures, examiners varied in their propensity to detect evasion (e.g., because of different experience). The DCE methodology we incorporate in Section 5 was designed to address this issue.

[14]Appendix Figure A6a depicts the fraction of tax units subject to an operational audit by fiscal year for the top 1%, 0.1% and 0.01% of reported AGI over time. Consistent with publicly available data, audit rates increased from the beginning of our sample period until about 2013 and then fell due to budget cuts. Public data on audit rates comes from the annual IRS Data Book, available at `https://www.irs.gov/statistics/soi-tax-stats-irs-data-book`.

[15]We note that the difference between the NRP tax gap and the operational audit tax gap in Figure 2 may not be statistically significant, due to sampling variation in the NRP (see Appendix Figure A1). However, a statistically significant difference is not necessary for the substantive point we make here. If operational audits of 10% or less of the top 0.01% each year uncovers half of all evasion in this group, which is less extreme than the lower bound scenario but still very conservative, the NRP point estimate should be twice the operational audit total. The NRP point estimate is sufficiently precise to rule this out with high confidence.

[16]Although Figure 2 provides an informative comparison between what is assessed in operational vs. NRP audits, it is not directly informative about the true level of the tax gap at the top, for two reasons. First, since only a small and selected fraction of the population is subject to an operational audit, one cannot infer the true level of tax evasion using operational audit data alone. Second, because in operational audit data we only observe reported (not corrected) income in the year of the audit, we rank tax units by reported income in Figure 2. When ranking by reported income, top earners mechanically have low evasion, since they are selected on high declared income.

FIGURE 2: TAX GAP: OPERATIONAL AUDITS VS. POPULATION-WEIGHTED RANDOM AUDITS



Notes: This figure compares noncompliance detected in operational audit data to population estimated noncompliance in the NRP, focusing on the top 10% of the distribution. Unlike in other figures, we rank individuals by reported income. We plot total evaded tax assessed as a fraction of total tax due in each bin, for both operational audits (pooling fiscal years 2007–2013) and population-weighted NRP audits (pooling fiscal years 2006–2013). Random audits uncover a very small amount of evasion in the top 0.01% by reported income. Operational audits represent a lower bound on total evasion because they only capture evasion of audited taxpayers. Yet they uncover substantially more evasion than the NRP population estimate in the top 0.01%.

3 Offshore Evasion

In 2008, the IRS and the U.S. Justice Department began an ambitious crackdown on offshore tax evasion, described in Johannesen et al. (2020). Key steps in this process included 1) legal actions against specific foreign banks, during which banks turned over records on Americans' concealed accounts, 2) the establishment in 2009 of Offshore Voluntary Disclosure programs, whereby taxpayers could disclose prior noncompliance and pay penalties but avoid criminal prosecution, and 3) the 2010 passage and 2014 implementation of the Foreign Accounts Tax Compliance Act. In this Section, we leverage the retrospective information created by this crackdown to study how accounting for offshore evasion modifies evasion detected in 2007, the year preceding the start of the crackdown.

3.1 Background and Data on Offshore Evasion

To first give some texture to our analysis, we gathered all public records from the Department of Justice website listing indictments, settlements, or convictions for offshore tax evasion in 2009-2021 (DOJ, 2021). These records contain descriptions of 178 cases of prosecuted offshore tax evasion;

naturally, these cases are selected and likely on the high-end in terms of noncompliance. A few results are worth noting. First, noncompliance in this sample is sophisticated. In 78% of the cases, offshore assets were held through at least one intermediate offshore structure (e.g., shell company or trust), and in 35% of the cases through multiple intertwined offshore structures.[17] Second, the amounts of unreported income are highly skewed, ranging from hundreds of thousands of dollars to $200 million over several years. Third, in nearly all cases, individuals were accused or convicted of failing to pay tax on the financial income generated by the assets in the account. Moreover, in about 30% of cases, the records suggest that pre-tax business profits were also diverted offshore and the corresponding income taxes under-paid. Last, 93% of prosecuted individuals were male.

We now present the IRS micro-data we use for our analysis of offshore evasion. We construct two lists of individuals who are likely to have been evading taxes on income from their offshore assets prior to the crackdown that started in 2008. These lists build on Johannesen et al. (2020), who found that the initial wave of enforcement in 2008 and 2009 caused a large increase in the reporting of offshore accounts and the associated financial income. The first list comprises participants in the Offshore Voluntary Disclosure (OVD) Program. We gathered data on all participants in OVD Programs from 2009 to 2015 and matched 50,020 OVD participants to their individual tax returns.[18] We refer to this sample as the *OVDP* sample.

The second list consists of individuals reporting that they own offshore assets by filing a Foreign Bank Account Report (FBAR) for the first time between 2009 and 2011. U.S. persons that are the beneficial owners of more than $10,000 in offshore financial wealth have been required to disclose this wealth to the government since the 1970s by filing an FBAR. We use only those first-time FBAR filers with U.S. addresses, disclosing an account in a tax haven.[19] Johannesen et al. (2020) show that the large majority of these taxpayers had been evading U.S. tax on these assets prior to disclosing them in response to enforcement.[20] We match 31,752 such taxpayers to their individual income tax returns. We refer to this sample as the *FBAR* sample. Individuals in this sample disclosed $124 billion in offshore wealth between 2009 and 2011. Total reported FBAR wealth was about $275 billion in 2011, indicating that a sizable share of all wealth reported on FBARs corre-

[17] Because descriptions of cases do not always detail how concealed offshore assets were held, these numbers are lower bounds for the frequency of sophisticated holdings of assets.

[18] This 50,020 figure does not include approximately 6,000 program participants who we could not match to their individual tax returns. About two-thirds of these were businesses. The rest did not file a tax return in the year we wished to analyze, e.g., because their participation in OVDP was too recent.

[19] We use the same list of tax havens as Johannesen et al. (2020): Anguilla, Antigua and Barbuda, Aruba, Bahamas, Bahrain, Belize, Bermuda, British Virgin Islands, Cayman Islands, Cook Islands, Cyprus, Dominica, Grenada, Guernsey, Hong Kong, Isle of Man, Jersey, Liechtenstein, Luxembourg, Malta, Mauritius, Monaco, Netherland Antilles, Panama, Singapore, St. Kitts and Nevis, St. Lucia Island, Switzerland, and Turks and Caicos Islands. Johannesen et al. (2020) constructed this list using countries mentioned in OECD (2000), plus a few other countries with strong banking secrecy. This is not an official definition used by the IRS, which has no official list of countries it considers tax havens.

[20] We note that this sample may contain some individuals who had an offshore account for legitimate reasons and were unaware of their FBAR filing obligation prior to increased enforcement, but we find it less plausible that very high-income individuals with accounts in havens were unaware of these obligations.

sponded to wealth in tax havens newly disclosed by filers with U.S. addresses. As we discuss below, our preferred estimate of total wealth concealed in tax havens by US persons is around $1 trillion in 2007 (Alstadaeter et al., 2018). This suggests that around 12 percent of all the wealth that was concealed in 2007 was disclosed by individuals in our FBAR sample in 2009–2011.

We rank each individual in these lists in the income distribution, using income data from the tax year after the disclosure of the offshore account. The results in Johannesen et al. (2020) suggest that this is the year in which individuals start complying with their tax obligations. We rank individuals by adjusted gross income (AGI) for simplicity. We report similar results with rankings based on other definitions of income.

3.2 Empirical Analysis of Offshore Evasion

We first show that NRP audits very seldom detected offshore tax evasion in the time period we study. As the NRP random sample (which is stratified) and the offshore lists contain disproportionately many observations of high-income individuals, there is some overlap between them—i.e., taxpayers who were randomly audited and also show up in our lists of likely offshore evaders. Before tax year 2009, FBAR compliance was not specifically examined in NRP audits. After auditors started examining FBAR compliance, 41 taxpayers were subject to a NRP audit before they appeared on our list of first-time FBAR filers with US addresses and haven accounts. Of these, fewer than 5 were flagged by NRP examiners for non-compliance with their FBAR obligation. Among all first-time FBAR filers, 187 taxpayers were audited in the NRP before disclosing an offshore account in 2009-2011, and again in fewer than 5 cases was any FBAR non-compliance detected. Likewise, of the 98 OVD participants whose FBAR compliance was examined in a random audit before entering OVD programs, fewer than 5 were flagged for FBAR non-compliance. In brief, random audits very rarely uncovered offshore evasion among taxpayers who were in fact likely to be evading.

Two caveats about our interpretation of these findings are worth noting. First, these figures are based on small samples. Second, taxpayers with offshore evasion detected in a random audit would no longer be non-compliant by the time we construct our lists of likely evaders. To alleviate these concerns, in Figure 3a we plot the rates of detected noncompliance over offshore wealth in the full NRP sample (not restricting to taxpayers who also appear in our lists of likely evaders). These rates are extremely small. In fact, they are much smaller than the fraction of taxpayers from each group of the income distribution that appear on our lists of likely evaders (shown in Figure 3b), a lower bound for the true fraction of offshore evaders. As Figure 3b shows, the probability to appear on our offshore lists reaches close to 7% in the top 0.01% of the income distribution.[21] Far more

[21] Appendix Figure A4 shows that our choice of income definition matters little for these estimates. We compare specifications using AGI, as in Figure 3b, with total positive income (which essentially re-ranks those with large business losses towards the top), and the sum of interest, dividend, and capital gains income (a rough proxy for wealth). The profile is similar across these three income definitions, though it is steepest for financial capital income, followed by positive income.

high-income individuals appear on our lists of offshore evaders than the (population-weighted) estimates from NRP audit data would suggest, even though these lists only cover a subset of all offshore evaders.

Figure 3c shows the distribution of offshore wealth by rank in the income distribution using wealth reported on FBARs in the FBAR sample. Of the $124 billion in offshore wealth disclosed, 21% belongs to the top 0.01% of the income distribution and 29% to the rest of the top 0.1%. This offshore wealth is far more concentrated than non-hidden wealth, of which about 7% belongs to the top 0.01% according to estimates from Saez and Zucman (2016), as updated in Saez and Zucman (2020).[22] The contrast confirms that findings in the previous figures were not simply driven by the overall concentration of wealth at the top of the distribution, but that concealed offshore wealth is *especially* concentrated at the top. Even the relatively small amount of FBAR wealth attributed to the bottom 90% of the income distribution (17% when ranking by AGI and 11% when ranking by total positive income) may actually belong to the top. Most FBAR wealth in the bottom 90% comes from a small number of large accounts; the median level of FBAR wealth for those in the bottom 50% of the distribution is around $200,000 and the mean is $2.5 million. If the true income of these large account holders could be observed, it would likely be near the top. To illustrate how much this might matter, Figure 3c shows the impact of reassigning FBAR wealth from the bottom 90% to the top 10% of the income distribution, in proportion with the FBAR wealth already attributed to the top 10%.

One caveat is that our samples of likely evaders are not necessarily representative of all individuals engaging in offshore evasion. Both the FBAR and OVDP samples comprise individuals who chose to disclose their offshore wealth via the OVDP or via a likely "quiet disclosure" (see Johannesen et al., 2020).We do not have direct evidence on the direction of the selection, but data from other countries suggests offshore wealth may be even more concentrated at the very top than we estimate. Alstadsaeter et al. (2019) use leaked data from HSBC Switzerland and estimate that 52% of offshore wealth in that bank was owned by taxpayers in the top 0.01% of the wealth distribution in Scandinavia.

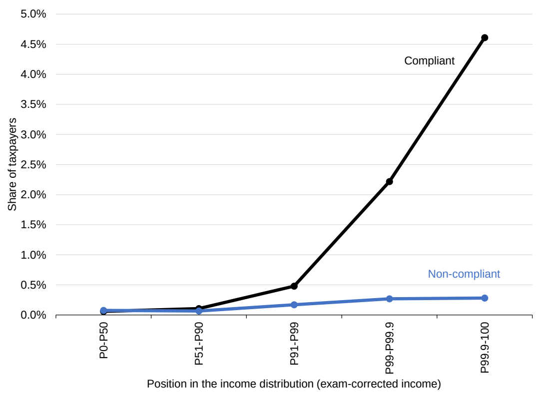
3.3 Implications of Offshore Evasion for the Distribution of Evasion

We now consider the magnitude of the adjustment to the income under-reporting gap implied by accounting for undetected offshore evasion. Following Alstadsaeter et al. (2019), we estimate the amount of unreported offshore income by proceeding in four steps. Each step entails an assumption. We summarize the assumptions we make in our preferred scenario and in sensitivity analysis in Appendix Table A3.

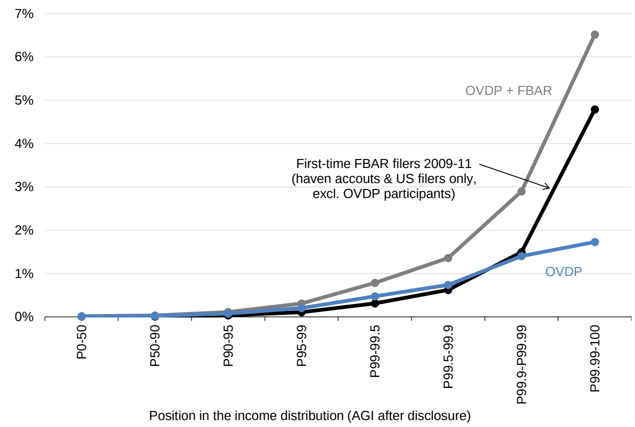
[22] Our focus is documenting the large difference between ownership shares of offshore and domestic wealth rather than the change in wealth shares over time. Using alternative estimates of wealth shares would lead to similar findings.

Figure 3: Findings from Data on Offshore Evasion

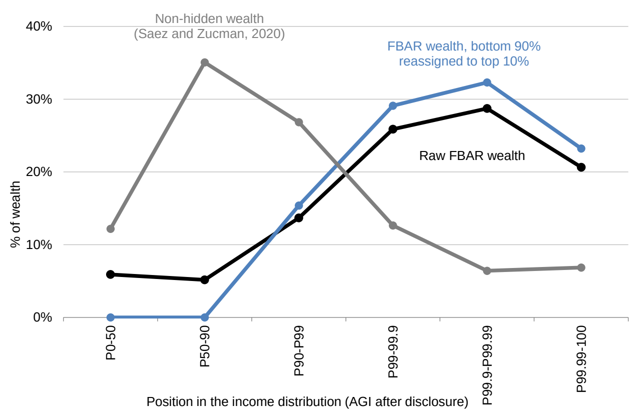
(a) Detection of offshore accounts in NRP



(b) Share Disclosing by Income Rank



(c) Distribution of Offshore Wealth



Note: This figure summarizes the main findings on offshore wealth from data on the OVDP and FBAR samples of likely offshore evaders. Panel (a) reports the share of taxpayers estimated to own either a compliant or non-compliant offshore account in the NRP random audit data. Auditors indicated whether each taxpayer had an FBAR filing requirement and whether the taxpayer complied with that requirement. We observe that rates of detected non-compliance are very low, much lower than the lower-bound rates of non-compliance shown in panel (b). Panel (b) plots the fraction of the full population that appears on our list of likely offshore evaders, by bin of adjusted gross income (measured in the tax year after disclosure of the offshore account), accounting for overlap between the lists (as OVD participants were required to file delinquent FBARs). We observe a steep profile of the probability of disclosing a previously hidden account by income rank. Nearly 7% of taxpayers in the top 0.01 percent of the income distribution appears in one of the two lists. In Panel (c), we plot wealth shares for non-hidden wealth from Saez and Zucman (2016) updated in Saez and Zucman (2020) by bin of market income (defined as total income reported on form 1040 minus Social Security benefits, unemployment insurance benefits, alimony, state refunds, and other income), versus wealth reported on FBARs by the first-time FBAR filers with U.S. addresses and accounts in tax havens (ranking by positive market income after disclosure). We observe that FBAR wealth is extremely concentrated at the top of the income distribution.

Benchmark Assumptions. In our benchmark scenario we make the following assumptions. First, we start with an estimate of aggregate offshore wealth in tax havens owned by U.S. households in 2007: $1,058 billion, the equivalent of 1.7% of total U.S. household wealth. This number is taken from Alstadaeter et al. (2018), Appendix Table A.3; it was obtained by Alstadaeter et al. (2018) by allocating the 2007 global amount of offshore wealth estimated in Zucman (2013) to each country, using retrospective statistics on the ownership of offshore bank deposits released by the Bank for International Settlements in 2016. Second, we assume that 95% of that wealth was hidden. Some accounts were certainly properly declared in 2007, but a 95% rate is consistent with the United States Senate (2008, 2014) reports, which found that 90%–95% of the wealth held by American clients of a number of Swiss banks were undeclared before FATCA.

Third, we assume a taxable rate of return on this offshore wealth of 6.0% to convert the stock of concealed offshore wealth to a taxable income flow. This rate of return is inferred from what is known about the portfolio composition of global offshore wealth and the rate of return on these assets in 2007. Specifically, Zucman (2013) estimates that in 2007, around 75% of global offshore wealth was invested in securities (mostly equities and mutual fund shares) and 25% in bank deposits. Our 6% taxable return is obtained by assigning the average interest rate paid by Swiss banks to deposits and half of the S&P 500 return to securities (with the other half consisting of unrealized capital gains).[23] We note that in a counterfactual policy environment where this wealth was not concealed and subject to tax, taxpayers might reallocate their portfolio toward assets generating unrealized capital gains and away from those generating taxable financial income. Tax gap estimates do not conventionally account for such behavioral responses. [24] On the other hand, we may under-state the proper conversion factor from offshore wealth to a taxable income flow because we account for financial capital income flows only, and not for evasion on the income that generated the principal in the offshore account. As discussed above, many offshore evaders diverted income from their business to their offshore accounts without paying tax to the US on the business income. The additional amount of evasion due to this issue could be significant. For example, if offshore wealth grew by 5% per year due to evasion on business income, then accounting for this would be equivalent to increasing our assumed "rate of return" by 5 percentage points.[25]

[23] The average interest rate paid by Swiss banks on their term deposits was 4.3% in December 2006. Specifically, at the end of 2006, 51% of the fiduciary deposits managed by Swiss banks were invested in US$, 29% in euros, and the rest in Swiss francs, British pounds, and yens (Swiss National Bank, 2007, Table 36). Money market interest rates were 5.4% for 3-months deposits in US$, 3.7% for 3-month deposit in euros, and 2.1% for 3-month deposits in Swiss francs (as reported in the money market rate statistics of the Swiss National Bank, `https://data.snb.ch/en/topics/ziredev/cube/zimoma`). The weighted average rate was thus 4.3% (taking the yield on 3-months francs as representative of the yield on deposits in currencies other than the euro and the US$). Moreover, in 2006 the U.S. Federal fund rate was in range of 4.3% to 5.25%; the total nominal return (dividends reinvested) was 13.4% for the the S&P 500 (and 20.65% for the MSCI world). With 25% of assets earning a 4.3% return in bank deposits and 75% earning half of a 13.4% return in securities (with the other half in unrealized capital gains), we arrive at our taxable 6% return.

[24] If newly disclosed assets were largely structured to generate primarily unrealized capital gain income, this could explain the relatively small implied taxable rate of return found by Johannesen et al. (2020), even if we suppose disclosed assets are primarily invested in equities.

[25] The 5% figure would be consistent with the offshore wealth being built steadily over about 14 years on average. If

Fourth, we distribute the aggregate amount of offshore wealth and income as follows. We take a weighted combination of the distribution of offshore wealth observed among self-selected U.S. filers who disclose a haven account by filing an FBAR for the first time in 2009-2011 (depicted in Figure 3c), and the distribution of hidden wealth estimated by Alstadsaeter et al. (2019) in Scandinavia. We put equal weight on the U.S. disclosed offshore wealth distribution and the Alstadsaeter et al. (2019) distributions. This implies that 60% of hidden wealth belongs to the top 0.1% and 35% to the top 0.01% (vs. more than 50% in Scandinavia).

Benchmark Estimates. Under the first three assumptions, unreported offshore income adds up to 0.7% of aggregate taxable income in 2007. As we saw in Section 2.2, before any correction for undetected evasion, the NRP finds that 4.0% of income is under-reported. Adding unreported offshore income increases this number to 4.7%. Figure 4a shows how adding offshore income modifies the estimated profile of evasion. Unsurprisingly, adding offshore income has no visible effect in the bottom 90% of the distribution and only a small effect between the 90$^{\text{th}}$ and 99$^{\text{th}}$ percentile. However, although offshore evasion is small on aggregate, accounting for it makes a significant difference at the top. It increases the ratio of under-reported income to true income by 4 percentage points in the top 0.01%, and by 3 percentage points for the rest of the top 0.1%. As a result, the sharp drop-off in the income under-reporting gap within the top 1% is undone by accounting for offshore evasion.

To estimate the consequences of offshore evasion for the tax gap, we must make a fifth assumption about the average marginal tax rate on income from offshore wealth. The average marginal tax rate should plausibly be between the top marginal tax rate on ordinary income and the preferred tax rate on long-term capital gains and qualified dividends, which are 35% and 15% in our reference year, respectively. Reflecting our earlier discussion about the portfolio composition of offshore wealth, we use an average marginal tax rate of 25% in our preferred scenario.[26] In total, we estimate that $15 billion in taxes was evaded from offshore accounts, with $10.5 billion of this total attributed to the top 0.1%, and $6.4 billion attributed to the top 0.01%.[27] With the lower 15% marginal tax rate, the total tax gap figure decreases from $15 billion to $9 billion, and with the higher 35% rate, it increases to $21 billion. Accounting for offshore evasion would therefore

we suppose wealth was built steadily over 30 years on average, our conversion factor is downward biased by about 2 percentage points. If we suppose the wealth was built steadily over 8 years or fewer on average, this implies a bias in excess of 10 percentage points. Finally, we note that the average marginal tax rate on the total income flow due to concealed offshore wealth would be larger if more of the income flow were business income.

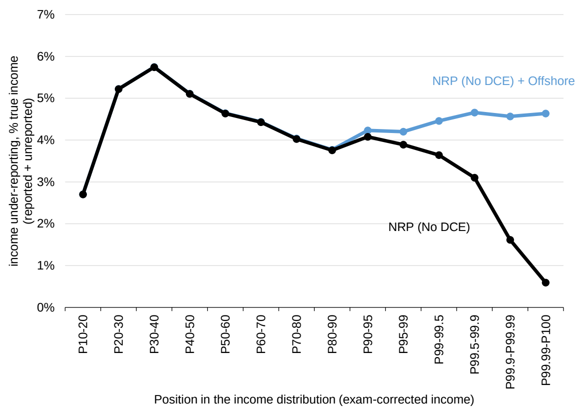
[26] This rate is consistent with a scenario in which 25% of taxable offshore income is interest income, 50% is long-term capital gains and qualified dividends, and 25% is short-term capital gains, non-qualified dividends and/or under-reporting of pre-tax business income diverted offshore; see also footnote 23.

[27] Our benchmark estimate of $15 billion in revenue loss is lower than the $23 billion estimate in Zucman (2014), because Zucman (2014) includes both federal and state taxes (and thus applies a 30% combined federal-plus-state marginal tax rate, as opposed to 25% in our benchmark scenario that captures federal taxes only) and assumes a 7% return (vs. 6% in our benchmark scenario). Zucman (2014) also estimates evaded estates tax (assuming 3% of offshore wealth belongs to decedents and a 40% estates tax rate), leading to total (income plus estate, federal plus state) tax evaded on offshore wealth of $36 billion (Zucman, 2014, Table 1 p. 140).

FIGURE 4: ACCOUNTING FOR UNDETECTED OFFSHORE FINANCIAL INCOME

(a) Unreported Income (% True Income)

(b) Sensitivity Analysis



Note: This figure plots the estimated income under-reporting rates with and without adding offshore tax evasion. Panel (a) shows our preferred scenario and panel (b) panel reports our sensitivity analysis. Taxpayers are ranked by exam-corrected market income in the NRP data, and offshore adjustments are made on the basis of positive market income (to proxy for "true income"). We find that income under-reporting rates increase significantly at the top of the income distribution when accounting for offshore evasion, reversing the drop-off in estimated evasion at the top seen in uncorrected random audit data. The point estimate for the top 0.01 percent increases by 4 percentage points in our benchmark scenario.

increase the tax gap at the top of the distribution significantly.

Sensitivity Analysis. We present results of our sensitivity analysis in Figure 4b. For simplicity, we focus on two scenarios: one in which each of our four assumptions is chosen (given the available evidence) to minimize the amount of offshore evasion at the very top, and one in which each assumption is chosen to maximize it – see Appendix Table A3 for a summary. These scenarios provide plausible lower and upper bounds for the size of offshore tax evasion at the top of the income distribution.

The lower bound of the amount of offshore wealth ($750 billion) comes from the Boston Consulting Group's (BCG) Wealth Report of 2007, which estimated that wealthy North American residents held about $37.7 trillion of wealth, 2% of which was held offshore.[28] The upper-bound is based on Guttentag and Avi-Yonah (2005), who built on the BCG Wealth Report of 2003, according to which the total holdings of high-net-worth individuals in the world were $38 trillion, including $16.2 trillion for North America residents. "Less than 10%" of this wealth was held offshore according to BCG; using this percentage as an upper bound, as in Guttentag and Avi-Yonah (2005), yields $1.5 trillion of U.S. offshore wealth. The lower-bound of the fraction of offshore wealth which is hidden is based on United States Senate (2008, 2014) reports investigating the practices of

[28] Alstadsaeter et al. (2019, footnote 28 p. 2090) list all the available estimates of the global amount of offshore around 2007; the BCG estimate is the second lowest one, immediately after (and close to) an OECD estimate which is not broken down by country.

several Swiss banks in the U.S. In these reports, the investigation committee find that about 90% of the wealth held by U.S. taxpayers at UBS Switzerland was undeclared, and that between 85% and 95% of the accounts held by U.S. taxpayers at Credit Suisse were undeclared. For the taxable rate of return, the conservative figure corresponds to the average daily 10-year Treasury rate for the year 2007 and is similar to the interest rate on Swiss bank deposits, a floor for the overall return since only about 25% of offshore wealth was invested in deposits. The upper-bound number is the return on average equity for all U.S. banks, averaged over the year 2007. Total income under-reporting via offshore accounts is $60.3 billion in the preferred scenario (0.7% of true total taxable income), $28.7 billion in the lower bound scenario (0.3% of total income), and $165 billion in the upper bound scenario (1.9% of total income). Finally, our preferred estimate of the distribution of offshore wealth and income was a weighted combination of our FBAR distribution and the distribution from leaks in the Nordic countries (Alstadsaeter et al., 2019). For the sensitivity analysis we put 100% of the weight on one or the other of these.

Two conclusions emerge from Figure 4b. First, there is uncertainty in estimates of unreported offshore income, which is reflected in the margin between the lower- and upper-bound aggregates above. In the upper-bound scenario, under-reported income as a share of true income is 9.7 percentage points higher than in our preferred scenario for the top 0.01%, while in the lower bound scenario is it 2.8 percentage points lower. Second, and interestingly, even in the lower-bound scenario in which concealed offshore income is much smaller (0.3% of taxable income), accounting for offshore evasion still has a large impact on estimated evasion at the top. It erases the downward-sloping profile of unreported income (as a fraction of true income) suggested by the NRP random audit data from the 99$^{\text{th}}$ percentile to the 99.99$^{\text{th}}$ percentile, while a drop-off remains in the top 0.01% in the lower bound scenario. In the lower bound scenario, accounting for offshore evasion also doubles the amount of tax evasion detected in the NRP for the top 0.01%. The striking and non-obvious result here is that even under very conservative assumptions, taking offshore evasion into account implies large adjustments to estimated evasion at the top.

In Appendix Figure A5, we unpack each step of the sensitivity analysis to see which assumptions matter most. Specifically, we modify assumptions from the preferred scenario one-by-one to arrive eventually at the upper and lower bound scenarios. We observe that the taxable return on offshore wealth is the most important source of uncertainty. Our own assessment is that the low rate of return used in the lower-bound scenario (4.5%, the 10-year Treasury yield) is likely too low for individuals in the top 0.01% of the income distribution in 2007, given that a large fraction of offshore wealth was invested in equities and evasion on pre-tax business income is likely significant on top of the financial income generated by offshore wealth. However, direct evidence on this question is limited. The next-most important assumption after the rate of return is the distribution of offshore assets. Changing this distribution primarily affects the amount of offshore evasion allocated to the top 0.01% versus the rest of the top 1%.

Finally, it is worth asking how our results, which are for the year 2007 (before the increase in enforcement effort on offshore wealth) can inform knowledge about top-end evasion post-crackdown. The available evidence suggests that post-2007 enforcement may have substantially reduced offshore evasion (Johannesen et al., 2020; De Simone et al., 2020). In particular, the implementation of the Foreign Account Tax Compliance Act in 2014 has significantly increased the information available to the IRS. We will take these facts into account in Section 6 when we present our preferred estimates of top-end evasion in the United States. In any case, these results are the first empirical demonstration in the U.S. context that some forms of evasion are highly concentrated at the top of the income distribution, effectively invisible in random audit data, and quantitatively important for the overall tax gap at the top. In what follows we provide evidence on another (possibly rising) form of evasion that shares these properties, namely tax evasion occurring via pass-through businesses.

4 Evasion on Pass-Through Business Income

Pass-through businesses (S corporations and partnerships) are not subject to the corporate income tax. Instead, all of the income of these businesses "passes through" to their owners' tax returns, where it is subject to applicable taxes. Ownership of pass-through entities is highly concentrated among the highest-income taxpayers, and the use of pass-through business structures has been on the rise since 1986 (Cooper et al., 2016). As such, obtaining accurate estimates of noncompliance in such structures is increasingly important. In this section we attempt to make progress on this question by leveraging additional data, focusing on the benchmark year 2007 to facilitate aggregation (in Section 6 below) with the offshore evasion results.

4.1 Background on Pass-Through Businesses

Administratively, pass-through businesses file returns reporting entity-level income. They allocate this income to their owners on a form called the Schedule K-1. Individual owners should report the income allocated to them on Schedules K-1 on their individual tax returns. When the owner of a pass-through business can control what the business reports, the Schedule K-1 issued by that business is not from an independent or unrelated third-party; in businesses where there is a single owner or a tight network of owners, the potential scope for noncompliance is similar to sole proprietorship income.[29] When the owner is passive, the business may under-report income and as a result allocate too little income to its owners. Confirming the flow of income, deductions, credits and other tax features of pass-through businesses takes expertise and resources. Partnerships create a specific additional challenge to the audit process, because partnerships can be owned by

[29]In 2017, according to Statistics of Income (SOI) tabulations of S corporation tax returns, 66% of all S corporations (earning 42% of all S corporation business income) had a single shareholder. When there is more than one owner, misreporting of pass-through income may be riskier than misreporting of sole proprietorship income (Kleven et al., 2016).

other entities, sometimes leading to complex ownership structures involving numerous partnerships, corporations, trusts, or other intermediaries (Cooper et al., 2016).

Conceptually, there are two types of potential non-compliance by individual owners of pass-through businesses: *individual-level under-reporting,* e.g., failure to report income allocated to the individual on a Schedule K-1, or mis-reporting whether that owner is active or passive (which has tax consequences), and *entity-level under-reporting*, when income on the pass-through business return is under-reported. In the context of NRP random audits of individuals, resource constraints, audit procedures, and the tools available to auditors limit the examination of entity-level under-reporting. First, as detailed below, when individuals report partnership or S corporation income on their individual tax returns, auditors rarely examine the tax returns of the corresponding pass-through businesses. Significant resources and expertise are necessary to audit up through pass-through returns and ensure that the correct income (and other tax features) flow through to the relevant individual income tax returns. The IRS does conduct audits to examine tax compliance of pass-through businesses and their owners, but these audit programs are unrelated to NRP random audits of individuals. Second, there is no recurring random audit program of pass-through business returns themselves. The most recent random audit program for partnerships was conducted in 1982. A small random audit program on S corporations was conducted for tax years 2003-2004.

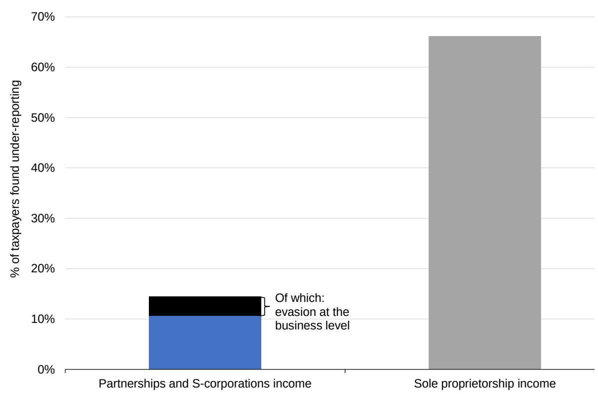
This situation has a number of implications. First, individual random audit data likely underestimate tax evasion occurring via entity-level under-reporting. In Appendix Table A1, we see that in these data, only 4.6% of partnership and S corporation income is found to be under-reported, compared to 36.7% for sole proprietorship income, which is subject to direct and comprehensive examination. Figure 5a contrasts the probabilities that noncompliance was detected in NRP individual random audits for sole proprietorship income (Schedule C) and pass-through business income (Schedule E). About 60% of the population of taxpayers with Schedule C income are estimated to be under-reporting their Schedule C income. By contrast, only 14.5% of taxpayers with pass-through business income are estimated to be under-reporting their pass-through business income. These cases primarily correspond to individual-level under-reporting rather than entity-level under-reporting. Only 3.8% of taxpayers with pass-through business income are found to be under-reporting income via an audit of the pass-through business return.[30] Detected entity-level under-reporting represents one-quarter of all instances of detected pass-through under-reporting.

Second, when entity-level reporting is examined, detected noncompliance tends to be high. As reported in Figure 5b, conditional on an entity being examined, more than 70% of business income is found under-reported on average, as opposed to 50% for sole proprietorships. Consequently, the

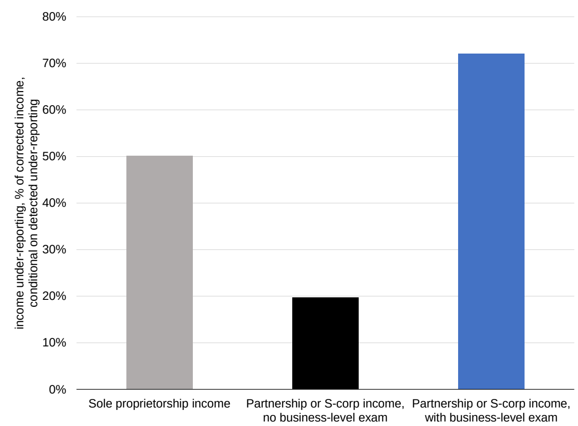
[30]More NRP related business audits occur for S corporations than for partnerships: About 75% of the audited related businesses are S corporations despite the fact that more NRP participants are partnership owners than are S corporation owners. Based on conversations with experts, our understanding is that when audits of pass-through businesses occur in the context of the NRP, the audited pass-through entities are typically small, simple businesses where the individual taxpayer being audited has access to the business's records. Large pass-through businesses—where income is concentrated—are very rarely examined.

FIGURE 5: PASS-THROUGH INCOME: DETECTED EVASION IN THE INDIVIDUAL RANDOM AUDITS AND CONCENTRATION

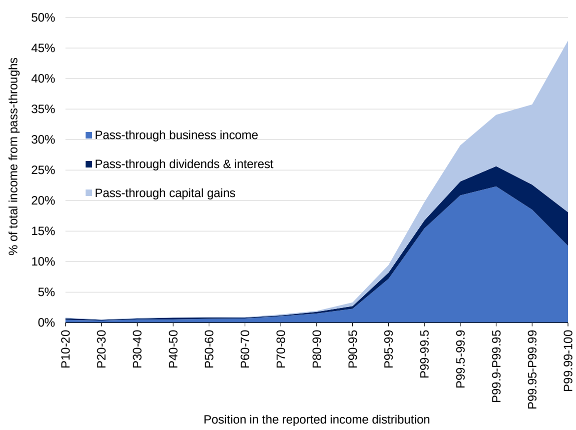
(a) Probability to Detect Unreported Income



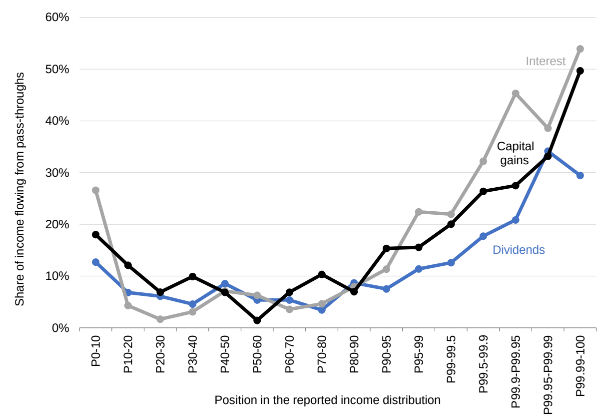
(b) Income Underreported (% True Income)



(c) Reported Income Earned via Passthroughs



(d) Investment Income Earned via Passthroughs



Note: Panel (a) shows the unconditional probabilities that noncompliance is detected for sole proprietorship income (schedule C) and pass-through business income (schedule E) in 2006–2013 NRP random audits. Only 14.5% of taxpayers with pass-through business income are found to be under-reporting their pass-through business income. Among these cases, only a quarter correspond to cases in which evasion is detected at the business level. Panel (b) shows the fraction of true business income found underreported in the in 2006–2013 NRP random audits for sole proprietorships, passthrough businesses when the entity is examined, and passthrough businesses when the entity is not examined. Panel (c) shows the fraction of reported market income earned via pass-through businesses by reported market income in 2007, by bin of reported market income. Income earned via pass-through businesses includes business income (S corporation and partnership profits), investment income (capital gains, dividends, and interest), and positive rents (lumped with business income) flowing from pass-throughs. For rents, we assume that 50% of positive rental income derives from pass-throughs and assume that this ratio is constant across the distribution; for other forms of investment income, pass-through investment income is taken from Panel (d). Panel (d) shows the fraction of reported dividends, interest, and capital gains earned via pass-throughs, by reported market income. These fractions are estimated in the population-weighted 2006–2013 NRP sample by matching 1040 individual income tax returns to Schedules K-1.

small number of cases where examiners found under-reporting at the entity level (one quarter of cases with detected under-reporting) account for 58% of all detected pass-through under-reporting in dollar terms. Naturally, the high under-reporting in examined businesses may partly reflect selection (auditors may be more likely to audit entities with observable traces of under-reporting). Still, this finding suggests that a lack of comprehensive examination of pass-through entities may bias overall estimates from individual random audit data significantly.

Third, because the ownership of pass-through businesses is concentrated, the bias occurs primarily at the top of the distribution. Figure 5c shows the fraction of reported income that the various groups of the reported income distribution earn via pass-through businesses in 2007, our benchmark year. For the bottom 90%, less than 5% of income comes from pass-throughs. However, among the highest earners this fraction increases to 45%. In other words, comprehensive audits of associated pass-through entities would be required to fully examine almost half the income reported by the highest-income individuals.

Fourth, the bias affects several income categories, not only partnership and S corporation business income. As shown by Figure 5c, the income that flows to individuals from pass-throughs does not only consist of business income, but also includes dividends, interest, rents, and capital gains. Determining compliance for such pass-through investment income (reported on Schedule K-1) is more resource-intensive than for investment income reported (on 1099 Forms) by third parties directly to the individual and the IRS.[31] On aggregate, about 15% of dividends, 20% of interest, and 30% of capital gains flow from pass-through businesses. The size of these flows (which largely go unexamined in the context of individual random audits) suggests potential for a significant underestimation of aggregate noncompliance on investment income. Moreover, as shown by Figure 5d, the fraction of investment income that derives from pass-throughs is highly skewed, so that the potential bias is again concentrated at the very top. In the top 0.1% of the income distribution, about 45% of interest and capital gains and 30% of dividends flow from pass-throughs.

Fifth, the magnitude of the tax gap due to pass-through evasion, and thus the magnitude of the bias from failing to account for entity-level pass-through evasion, is likely to have grown over time. Appendix Figure A7 illustrates how partnership and S corporation income has grown as a share of overall income at the top of the distribution since the 1980s. The first panel shows that top 1% income shares have risen substantially over time (by 1.83 times), and the majority of that growth (66%) is attributable to pass-through income, which rises from 11% to 36% of top 1% incomes over this period. The second panel isolates the trend in partnership and S corporation income going to the top 1% as a share of total income, which rises steadily over this period (from 1% to 7%), including through the period of heightened offshore enforcement. Additionally, Figure A6b

[31] A further complication is that income that should be taxed as e.g. business income may be reported as capital gains due to the lower tax rate on long-term capital gains. There is potential for legal avoidance along these lines, noncompliance and, possibly, some gray area.

depicts audit rates of individuals, pass-throughs and other corporations over time using data from IRS (2022). We observe that pass-throughs were already subject to low audit rates prior to the precipitous decline in the last decade. These already low audit rates declined by over 80% from 2011 to 2019; the current audit rate for partnerships and S-corporations is about 0.05%, which is lower than the audit rate for even very low-income individuals and very small C corporations.[32] We do not have data with which to credibly quantify the growth in evasion via pass-throughs over time, but these two facts suggest that, perhaps unlike offshore evasion, pass-through evasion may have grown substantially in recent years.

4.2 Evidence from Random Audits of S corporations

To provide quantitative insights on entity-level under-reporting, we analyze data on the S corporation random audit program from tax years 2003-2004. The data come from audits of 4,515 S corporation tax returns. For line items on the S corporation tax return (Form 1120-S), we observe originally reported and exam-corrected amounts.[33] The sample is stratified based on risk and assets; large S corporations are over-sampled (for more details, see Johns, 2009). We use sampling weights to construct representative population estimates, pooling the years 2003 and 2004.

We estimate entity-level under-reporting as under-reported receipts plus over-reported deductions. Following Johns (2009), we do not include corrections to officer compensation deductions in our measure of overall income under-reporting because those corrections are offset one-for-one at the individual owner level by changes in wage income.[34] To examine entity-level under-reporting through the individual income distribution, we also linked the audited S corporations' returns to their owners' tax returns via Schedules K-1. To obtain the best available estimate for owners' rank in the true income distribution, we use "quasi-corrected" income, defined as reported (market) income from the owner's Form 1040 plus the individual owner's share of detected entity-level S corporation under-reporting (excluding mis-reported officer compensation).

These data are subject to the same potential limitations as individual random audit data. First, we make no corrections for undetected evasion, as in Section 2.2. Second, resource constraints and audit procedures can limit the comprehensiveness of the examination of noncompliance, especially in large businesses. Like individual random audits, these S corporation audits were conducted

[32] In 2019, individuals with between $1 and $25,000 of total positive income were subject to an audit rate of 0.37%, while C corporations with between $1 and $250,000 of assets were subject to an audit rate of 0.16%. For further details refer to Table 17 of IRS (2022). The audit rate for partnerships and/or S corporations in 2019 is the lowest of any group of tax returns for which an audit rate is reported in Table 17. In earlier years, the audit rate for partnerships and S corporations is among the lowest of the group-specific audit rates reported in the table.

[33] The focus of the audit program was on the accuracy of business income reporting, i.e., the front page of the Form 1120-S. We do not have data on examinations of, e.g., tax credits claimed by S corporations (i.e., the information on Schedule K of the 1120-S).

[34] For example, an owner who chooses to under-report officer compensation relative to the "reasonable compensation" standard will report too much ordinary business income relative to that standard. Mis-classification of officer compensation as ordinary business income allows owners to avoid Medicare and potentially other employment taxes. This type of non-compliance is frequently detected in these audits (Johns, 2009). It matters for the tax gap but not for the extent of income under-reporting.

by examiners from the Small Business and Self-Employed (SBSE) division of IRS, and the audit procedures were the same as routine SBSE audits of S corporations. Given these limitations, we interpret the data as likely providing a lower bound for the extent of entity-level under-reporting in pass-through businesses. In distributing overall non-compliance through the individual income distribution, we face a third limitation: we do not observe non-compliance on the individual tax returns (Forms 1040) of the owners of the audited S corporations. Insofar as owners of under-reporting S corporations tend also to under-report elsewhere on their individual tax return (see e.g. Joulfaian, 2000), the estimated location of entity-level under-reporting will be biased toward the bottom of the distribution, due to insufficient re-ranking (see Figure A3 and Table A2 for an illustration of how accounting for re-ranking shifts under-reporting upward in the distribution).

We estimate that S corporations under-report 19.5% of true net business income. Appendix Table A4 presents a line-by-line decomposition. Of total under-reporting, 26% represents under-reported gross receipts or sales,[35] 69% over-reported deductions,[36] and 5% mis-reporting of net gains from the sale of business property (Form 4797) or "Other Income" (Form 1120-S line 5).

In Figure 6, we examine pass-through under-reporting through the individual income distribution. Figure 6a reports the distribution of 1) reported, 2) exam-corrected, and 3) under-reported S corporation income by owner income bin.[37] Both reported S corporation income and S corporation income corrected for entity-level under-reporting are concentrated at the top of the individual income distribution, with 49% (45%) of reported (corrected) S corporation income belonging to individuals in the top 0.1%. In total, 79% (75%) of reported (corrected) S corporation income belongs to owners in the the top 1%. The reported income shares are larger than the corrected income shares because we estimate that under-reported income is less concentrated than reported income, with 51% of under-reporting in the top 1%. Most of this difference is attributable to a drop-off in estimated mis-reporting rates at the very top of the owner income distribution, which we further examine in Figure 6b.

Figure 6b reports income under-reporting rates in different parts of the owner income distribution, separately for large S corporations—defined as those in the top decile of the net assets distribution—and others. For S corporations in the bottom 90% of the asset distribution, the decline in under-reporting rates as one moves up the income distribution is not entirely surprising given the characteristics of many of these firms (see, e.g., Smith et al., 2019b). In P90-P99.5, S corporations often comprise self-employed, skilled professionals operating as S corporations. These

[35] The under-reporting rate for receipts on their own is lower than most other line items: 0.3% of gross receipts is under-reported, but 0.3% under-reporting rate of gross receipts translates to 5.1% of net S corporation income.

[36] Over-reported deductions here excludes officer compensation but including the cost of goods sold. The largest component of over-reported deductions is Form 1120-S Line 19: "Other Deductions." However, some corrections to line 19 pertain to line-switching, i.e. when an examiner determines that deductions reported on line 19 should have been entered elsewhere. Such line switching will net out for our primary purpose here, but it necessitates caution when interpreting Table A4 in terms of which types of deductions are most commonly under-reported.

[37] We are unable to use the finer income bins we use in previous figures due to the smaller sample size of the S corporation study.

corporations are not very different from sole proprietorships, and indeed we estimate comparably high under-reporting rates of about 30%. In the top 0.1%, small-asset S corporations are more likely to represent supplemental income to the individual's main sources of income, e.g., speaking fees or consulting work, where there may be more information reporting and less evasion. This decline in the under-reporting rate may be over-stated due to the insufficient re-ranking discussed above, however.

S corporations in the top 10% of the assets distribution are more similar to C corporations. For these large S corporations, we estimate under-reporting rates of about 15% through most of the income distribution, similar to the C corporations under-reporting rate suggested by IRS tax gap studies of 18% IRS (2016a).[38] For top 0.1% owners, however, the under-reporting rate for large S corporations falls to about 5%. There are at least two possible interpretations for this finding. First, there might truly be little noncompliance in these large, top-owned private firms. Second, there may be significant noncompliance in these large private firms that the random audit program failed to uncover. This may be a consequence of audit procedures: thoroughly auditing multi-establishment S corporations (e.g., retail chains or car dealerships), which are concentrated in the top 10% by asset $\times$ top 0.1% by income group, requires far more resources than auditing a small S corporation. Given this, we approach with caution the low estimated rate of entity-level S corporation under-reporting at the very top of the individual income distribution, which echoes the drop in detected individual noncompliance in Figure 1a.

To quantitatively investigate this issue, we analyze data on the participation in two specific evasive schemes involving pass-through entities: microcaptive insurance schemes and syndicated conservation easements. In evasive micro-captive insurance schemes, business owners pay a fee to an offshore company that they also own and supposedly provides insurance services. This payment is deducted as a business expense in the US, reducing tax liabilities there, while the fee does not increase tax liabilities because the offshore company that receives it is incorporated in a territory where there is no income tax (e.g., Bermuda). If the micro-captives do not meet the legal criteria for providing insurance, then the payments are evasive (for more details, see IRS, 2016b). The IRS identified microcaptive schemes as concerning and has taken steps to curb them, including a letter campaign and the issuance of Notice 2016-66, which identified the schemes as potential tax evasion and required taxpayers who had participated in these schemes to disclose their participation on Form 8886. To identify participants in microcaptive insurance schemes, we use the initial Form 8886 filings in response to IRS Notice 2016-66. We observe 12,874 S corporation owners disclosing participation in microcaptive insurance schemes during tax years 2017 to 2020 and we assign them to income ranks using their AGI in the year after Form 8886 filing.

[38]The gross corporate tax gap is estimated at $44 billion on average in tax years 2008-2010 (IRS, 2016a, p. 7), a period during which corporate income tax revenues averaged $191 billion (National Income and Product Accounts, Table 3.2 line 8), hence a tax gap of 44 / (191 + 44) = 18.8% of true income.

Syndicated conservation easement schemes involve the donation of the development rights on property to a non-profit entity for conservation, which generates a tax deduction. By overvaluing the development rights, taxpayers in evasive syndicated conservation easement schemes claim deductible donations larger than the deductions they are entitled to. The word "syndicated" refers to the version of the conservation easement where a group of individuals jointly own a pass-through business (typically a partnership) that owns the property on which the conservation easement is taken (for more details, see United States Senate, 2020).[39] Thus the syndicated conservation easement is another sophisticated evasion scheme involving pass-through businesses, though it involves income tax deductions (i.e., on Schedule K of the entity return) rather than net business income. To identify participants in syndicated conservation easements, we use the initial Form 8886 filings in response to IRS Notice 2017-10, which identified syndicated conservation easements as "transactions of Interest." We observe 26,117 such 8886 filers during tax years 2017 to 2020 and we assign them to income ranks using their AGI in the year after 8886 filing.[40]

Figure 6c presents the probability that an individual appears on our list of S corporation owners participating in a micro-captive insurance scheme by individual income bin. To help us understand how these findings relate to the S corporation random audit results, we report separate probabilities of micro-captive participation in large and small S corporations, splitting out firms in the top decile vs. bottom 90% of the assets distribution as before. Figure 6d plots the probability for participation in a syndicated conservation easement transaction. In both of these figures, as for offshore evasion in Figure 3b, we observe that participation in sophisticated schemes rises sharply with income even within the top 1%. For micro-captive insurance, we further find that the relatively high participation rate among very high-income owners is almost entirely driven by large S corporations, which contrasts sharply with the random audit findings in Figure 6b. Figures 6c and 6d suggest that some of the top-end drop in detected entity-level under-reporting in S corporations from Figures 6a and 6b reflects differential detection of evasion through the distribution of S corporation size $\times$ owner income. The available data, however, limit our ability to definitively correct for the under-detection of sophisticated, undetected, entity-level under-reporting in passthrough businesses. Regarding partnerships, we also note that the complexity of partnership

[39] There are also evasive non-syndicated conservation easements, but these are somewhat rarer and their relationship with pass-through compliance specifically less direct.

[40] Both firms and individuals respond to these notices. For microcaptive insurance schemes, we limit our analysis to S Corporation owners to maximize comparability with the S Corporation random audit data. For syndicated conservation easements, S corporation owners are less well-represented, so we include individuals filing Form 8886 and also the individual owners of the passthrough businesses filing Form 8886. In the case of passthrough ownership, there is some subjectivity about which owners should be included. First, some owners may not even claim the relevant charitable deduction, so we restrict our sample to owners who do claim the Schedule A line item in the year of interest or a recent prior year. Second, some owners, e.g., minority investors in a complex fund of funds, may not know they are invested in the position. Together with the Schedule A line item restriction, we use the materiality of the owner's stake in firms engaged in this position to proxy for the likelihood that the owner knows about the position. We require the owner owns at least 0.01% of the pass-through; we obtain a very similar profile with a 0.1% ownership or 1% ownership restriction, or with no such restriction.

structures (e.g., the use of tiered structures) rises sharply with income within the top 1% of the income distribution, as shown in Appendix Figure A8. Given all this, we regard the magnitude and concentration of under-reported S corporation income in Figure 6a as a lower bound for the magnitude and concentration of overall entity-level under-reporting in S corporations and partnerships, and illustrate a range of possibilities for overall pass-through evasion and its distribution in the next section.

4.3 Implications of Pass-Through Evasion for the Distribution of Evasion

In this section, we assess the magnitude of the bias in the individual random audit data from Figure 1a due to undetected pass-through business evasion.

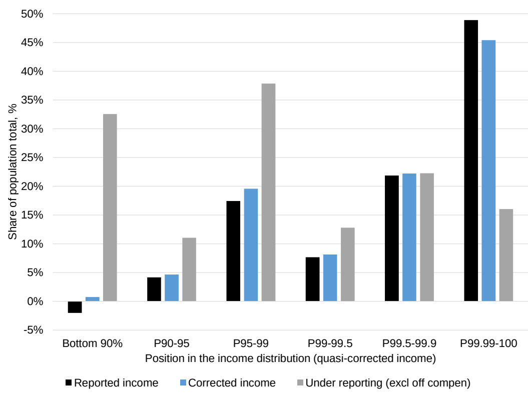
From a macro perspective, the key questions governing this bias are 1) total amount of entity-level under-reporting for different types of pass-through income, 2) the shares of this under-reporting belonging to owners at different parts of the individual income distribution, and 3) the amount of entity-level under-reporting that is already detected in individual random audits. We illustrate the magnitude of the bias in a benchmark scenario wherein we make empirically grounded assumptions about all three of these quantities, and then we illustrate alternative scenarios in sensitivity analysis.

Assumptions. In our benchmark scenario, we assume that 20% of total pass-through business income, 5% of pass-through capital gains, and 3% of pass-through dividends and interest are under-reported. The under-reporting rates assumed for pass-through investment income equal the under-reporting rates for investment income directly earned by individuals, as detected in the individual random audit data (see Appendix Table A1). For business income, our benchmark 20% under-reporting rate is motivated primarily by the estimate from the S corporation random audit data discussed above. We argue that this is a conservative figure for under-reporting of all pass-through business income, in partnerships and S corporations, because this figure does not account for undetected evasion in the S corporation random audit study and because the complexity of partnership arrangements generally create more scope for non-compliance in partnerships than in S corporations. The last random audit study of partnerships happened in 1982, when TCMP random audits produced an estimated income under-reporting rate of 26% for partnerships, though obviously in a different tax policy regime (GAO, 1995). Our benchmark 20% figure also lies between the random audit figure for sole proprietorships (37% under-reporting rate) and the tax gap figure for C corporations (18%), which is reasonable given the nature of most pass-through business activity.

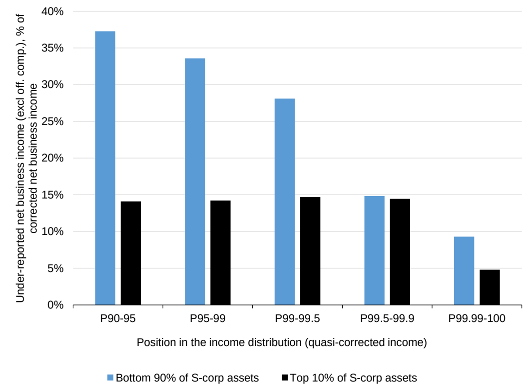
We also suppose in the benchmark scenario that undetected pass-through income is distributed like reported pass-through income. For example, if the top 1% (by reported income) earns 60% of reported pass-through business income, we assume that the top 1% (by true income) also earns 60% of unreported pass-through business income. Given what we infer from Figure 6 about the

FIGURE 6: PASSTHROUGH EVASION: S-CORPORATION RANDOM AUDITS & SPECIFIC SCHEMES

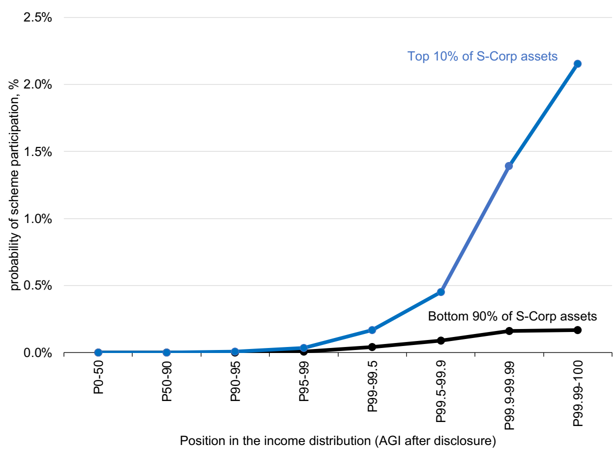
(a) Owner-level concentration of reported, corrected, & under-reported business income



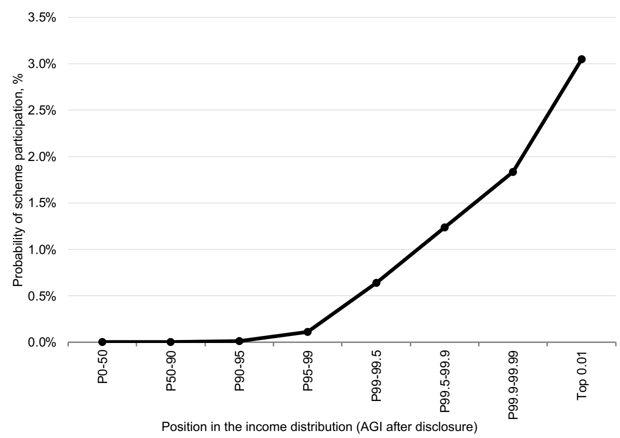
(b) Rates of entity-level under-reporting



(c) Micro-captive participation rates



(d) Syndicated conservation easement participation rates



Note: This figure presents data on entity-level under-reporting in pass-through businesses. Panels (a) and (b) are drawn from the S corporation random audit data from tax years 2003–2004. We rank individuals by their reported income plus the S corporation owners' share of S corporation under-reporting, which we call "quasi-corrected" income. Panel (a) presents the distribution of reported, exam-corrected, and under-reported S corporation net business income by owner income bins. We exclude mis-classified officer compensation from under-reporting for the reasons discussed in the text. Panel (b) depicts estimated rates of entity-level under-reporting in S corporations, by owner income bin, splitting large and small S corporations (i.e. top 10% versus bottom 90% by assets). We observe that under-reporting is highly concentrated but not quite as concentrated as reported business income, because of a drop in detected under-reporting for large S corporations with top-income owners. The next two panels show, however, that sophisticated evasion schemes are relatively frequently taken up by top-income owners of pass-through businesses. In panel (c) we depict the rate of participation by owners of S corporations in micro-captive insurance schemes. We split these owners into two groups based on whether they own small or large S corporations. We observe that sophisticated evasion via micro-captives is concentrated among top-income owners of large S corporations (the same group Panel (b) suggests are relatively compliant). In panel (d), we show that rates of participation in syndicated conservation easements are also high among top-income owners of partnerships and S corporations.

limitations of the S corporation random audit data in covering sophisticated evasion by top-income owners, assuming that the rate of under-reporting of business income is constant over the owner income distribution seems like a reasonable benchmark.

Finally, we add pass-through evasion to the amount of under-reported income detected in individual random audits by income bin. To avoid any double counting with noncompliance detected in individual random audits, we remove all entity-level pass-through evasion uncovered in the context of individual random audits.[41]

For our sensitivity analysis, we illustrate sensitivity around a few more specific issues involving partnerships, and we consider a lower-bound scenario and a high-end scenario. The lower-bound scenario presumes that entity-level financial capital income under-reporting is non-existant, while the S corporation random audit data from Figure 6 provide a complete and representative picture of all pass-through evasion, i.e. the rate of under-reporting is 19.5%, which is distributed according to the shares of S corporation under-reporting by owner income bin from Figure 6a. For the high-end scenario, we suppose 28% of pass-through business income is under-reporting (as suggested by the 1982 TCMP study on partnerships), 10% of pass-through capital gains, and 6% of pass-through dividends and interest are unreported. The higher rate on pass-through business income is motivated by the concern that the S corporation random audits may not detected a significant amount of evasion, and partnerships may be more non-compliant than S corporations. The higher rate on pass-through investment income is motivated by the fact that pass-through investment income is more sophisticated than the individual investment income on which these benchmark rates are based.[42] We distribute these larger aggregate totals for under-reported income like reported pass-through income, as in the benchmark scenario.[43] In both of these scenarios, we remove entity-level pass-through evasion uncovered in individual NRP audits to prevent double counting, exactly as in the benchmark scenario.

Benchmark Estimates. Figure 7a shows the results obtained with our benchmark assumptions. On aggregate, the pass-through adjustment (1.5% of true income) is about twice as large as the offshore adjustment (0.7%). Accounting for pass-through businesses increases under-reported in-

[41] Specifically, 57.6% of partnership and S corporation income evasion detected in individual random audits is associated with an entity pick up (i.e., an audit of the corresponding business). Therefore we remove 57.6% of detected partnership and S corporation evasion.

[42] For some assets owned by pass-through businesses—such as real estate, business structures and equipment, and foreign securities—the IRS does not know the purchase price of the assets, which enables tax evasion on capital gains. Interest and dividend payments flowing through pass-throughs often involve foreign entities, and especially before the implementation of the Foreign Account Tax Compliance Act, there was also little reporting on this type of income. We are conservative with our benchmark assumptions on under-reported investment income in the pass-throughs in part because some of this income may involve offshore entities, which creates double counting issues when we combine the pass-through and offshore adjustments in the next section. Remaining conservative on this point in the pass-through benchmark obviates the need to adjust further for such double counting below. Another source of uncertainty owes to the fact that taxpayers may attempt to classify income as capital gains to benefit from the relatively low rate on long-term capital gains (see footnote 31).

[43] Our approach does not factor in any re-ranking from adding undetected pass-through business evasion to reported income, which tends to make our estimates conservative with respect to the concentration of evasion at the top.

come at the top of the income distribution significantly. In estimates from unadjusted individual random audit data, the fraction of true income which is under-reported falls from 4% around the 90th percentile of the (exam-corrected) income distribution to less than 1% in the top 0.01%. After adding pass-through under-reported income, the fraction of true income which is under-reported rises from 4% around the 90th percentile to about 8% from the 99.5th to the 99.95th percentile. It then falls back to around 5% in the top 0.01%. This drop-off is partly due to the increasing prevalence of capital gains (as opposed to business income) at the very top; we assume pass-through capital gains have a lower under-reporting rate (5%) in our benchmark scenario. Tax evasion below the 90th percentile is not significantly affected by the pass-through correction, because pass-through income is a negligible source of income for these groups.

To further assess the effect of accounting for pass-through evasion, it is useful to consider the top 1% as a whole. When including only the pass-through evasion uncovered in the individual random audit data, the top 1% under-reports 2.3% of its true income—a lower rate than the average under-reporting rate of 4.0%. After accounting for pass-through evasion using our benchmark assumptions, the top 1% under-reports 6.7% of its income—a higher rate than the average rate of 5.3%. In other words, the pass-through correction alone removes the decreasing pattern of under-reporting rates found in uncorrected random audit data.

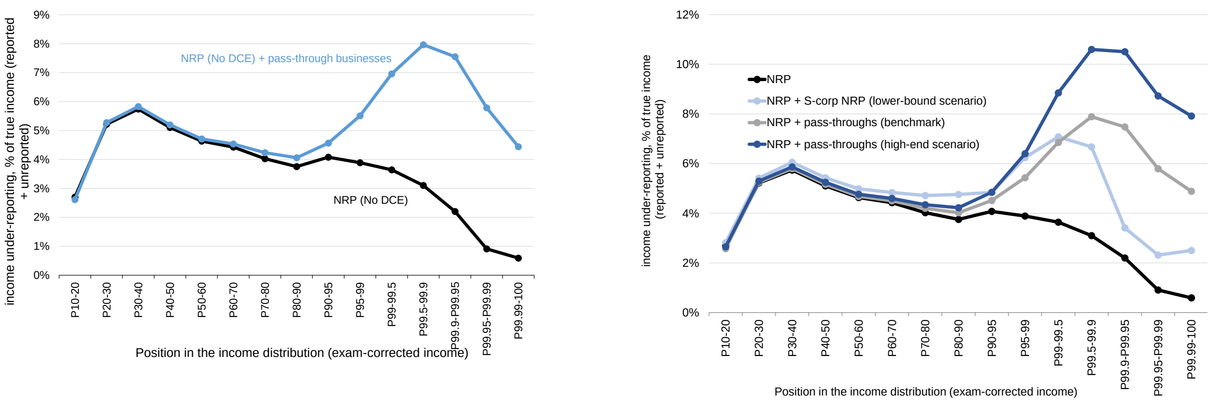
Accounting for pass-through evasion also increases the tax gap, by $34 billion on average over 2008–2013. The pass-through adjustment is more than twice as large as the offshore adjustment ($15 billion), reflecting the fact that in our preferred scenarios (i) unreported passthrough income is about twice as large as unreported offshore income, and (ii) the marginal rate on pass-through business income is slightly higher than the assumed marginal rate on offshore income (25%).[44] One-third of unpaid pass-trough business income taxes are attributable to the top 0.1%.

Sensitivity Analysis. Figure 7b illustrates our main sensitivity analysis, including the benchmark scenario, a lower bound scenario based on the S corporation random audit data on business income under-reporting and zero mis-reporting of pass-through investment income, and a high-end scenario with the higher mis-reporting rates described above. In the lower bound scenario based on the S corporation random audit data, top 1% evasion is more than twice what is detected in individual random audits. Relative to the benchmark, the somewhat lower concentration of under-reporting estimated from this data results in less evasion in the top 0.1% and more evasion in the bottom of the top 1% and the rest of the top 5%, which is consistent with Figure 6b. In the high-end scenario, mis-reporting rates are higher throughout the top 1% by about 2 percentage points. In Figures A9 and A10, we break down the sensitivity analysis in Figure 7b, illustrating sensitivity to our assumptions around pass-through business income while holding our assumptions pass-through investment income constant, and vice versa.

[44] For business income, which accounts for the bulk of unreported passthrough income, the marginal tax rate is the ordinary income tax rate, 35% in the top tax bracket.

FIGURE 7: ACCOUNTING FOR PASS-THROUGH BUSINESS EVASION

(a) Unreported Income (% True Income) (b) Sensitivity Analysis



Note: This figure shows estimates of unreported income by income group in the raw NRP and after adding estimates of pass-through business evasion. Taxpayers are ranked by exam-corrected income in NRP data, and pass-through adjustments are made on the basis of reported market income; this is the best available estimate of "true income". In our benchmark scenario (panel a), we assume that 20% of pass-through business income, 5% of pass-through capital gains, and 3% of pass-through interest and dividends are under-reported, and that under-reported pass-through income is distributed like duly reported pass-through income. We remove all business-level pass-through evasion detected in the NRP before adding our estimates of business-level pass-through evasion. In panel b, we report a high-end scenario in which 28% of pass-through business income, 10% of pass-through capital gains, and 6% of pass-through dividends and interest are unreported, and a lower bound scenario in which under-reporting of business income in both S corporations and partnerships follows the estimated rate of under-reporting and distribution from the S corporation random audit data from Figure 6, while all pass-through investment income is duly declared.

In the Appendix, we also consider a number of specific issues that could cause the benchmark scenario to under-estimate evasion. First, we randomly disallow 20% of declared pass-through business losses, setting income to zero instead of the reported loss. A significant share of adjustments to partnership income in operational audits consists of disallowed losses, which thus appear to be correlated with noncompliance. Appendix Figure A11 shows that this modification adds about one percentage point to rate of income under-reporting at the top. The effect of this modification is concentrated at the top because of re-ranking: taxpayers whose losses are disallowed typically are in the bottom decile of the reported income distribution, but end up in the top 1% of the corrected income distribution after their declared losses are set to zero. In a different modification, we classify part of the business income earned by circular partnerships—e.g. partnership A is a partner in partnership B, B a partner in C, which in turn is a partner in A—as tax evasion. Cooper et al. (2016) find that 15% of partnership business income is earned in circularly-owned partnerships in 2011; Appendix Figure A12 illustrates a scenario in which we classify two-thirds of this income (i.e., 10% of all partnership business income) as evasion. Doing so adds 0.7 percentage points to the ratio of under-reported income to true income in the top 1%. Third, we show in Appendix Figure A8 that the complexity of partnership arrangements increases with owners' income.

About two-thirds of top 0.01% partnership owners receive income from tiered partnerships, i.e. partnerships that are owned by other partnerships. More complex partnership arrangements create scope for some types of sophisticated evasion and grey-area avoidance. We have limited data to discipline an additional sensitivity test along these lines, but the stark difference in complexity of partnership income within the top 1% of the distribution suggests our current benchmark is conservative. In summary, adjusting for any of these issues will tend to increase estimated evasion primarily at the very top of the distribution, because of the concentration of pass-through income, rising complexity at the top, and the fact that disallowed losses lead to substantial re-ranking.

5 Incorporating Evasion Identified by Detection-Controlled Estimation

The previous two sections showed that offshore tax evasion and entity-level pass-through evasion were virtually never detected in individual random audits and quantitatively important. Our next objective is to understand what these forms of evasion imply for unreported income in general, compared to official statistics. To do this, we must first reconcile the random-audit-based results in Section 2.2 with official estimates of income under-reporting and the tax gap. Official tax gap statistics account for some types of undetected evasion using detection-controlled estimation (DCE) methods (IRS, 2019). Here, we augment our estimates of the profile of evasion to account for the forms of undetected evasion estimated by DCE. Doing so allows us to reconcile our aggregate estimates with official estimates of the federal individual income tax gap and to account for other types of undetected evasion besides the ones we considered in Sections 3 and 4. In this Section we present distributional estimates incorporating DCE-identified evasion *without* the forms of evasion from Sections 3 and 4; in Section 6 we combine everything.

5.1 Conceptual Overview of DCE and Methodology

Overview of DCE. The basic idea of DCE is to identify undetected evasion by modelling the probability of detection of under-reporting, and then to construct a counterfactual scenario where this probability equals one everywhere. Methods currently used by the IRS implement this idea using a model of auditor effects. The key underlying assumptions behind the method are 1) that auditor assignment rules generate exogenous variation in the probability of detection of under-reporting, i.e., auditors are assigned quasi-randomly,[45] and 2) there is a subset of auditors with 100% detection rates. Building on these assumptions, researchers estimate a model of the detection process to compute total evasion in a counterfactual where all auditors are replaced by those auditors with the 100% detection rates (for details, see Feinstein, 1991; Erard and Feinstein, 2011).

The evidence in Sections 3 and 4 suggests that the assumption that the top auditors have 100% detection rates is too strong, especially at the top of the distribution. In particular, Figures 3a/3b and 5a imply that the detection of offshore evasion and entity-level evasion in pass-through busi-

[45] A concern with this assumption is that auditors may be assigned to cases in a manner that selects on suspected evasion. We discuss this in the final section of Guyton et al. (2023).

nesses is so rare that these forms of evasion are unlikely to be detected even by thorough auditors (e.g., because they may not even attempt to do so under current audit procedures, as in the case of businesses-level evasion, or because they may not have the necessary information to detect evasion, as in the case of offshore accounts before FATCA). Nevertheless, if we relax the 100% detection assumption to account for the possibility that some under-reporting is undetectable by any auditor (due to audit procedures, information constraints, or resource constraints), DCE methods would still identify some undetected under-reporting. We can think of the forms of evasion identified by DCE methods as *detectable but undetected* evasion, which could include many forms of evasion besides the two forms of *undetectable* evasion we considered above.[46]

The main challenge in incorporating evasion implied by DCE into our analysis is that existing methods are primarily used to identify a total amount of undetected evasion, rather than evasion throughout the income distribution, especially the very top. Distributing the undetected evasion identified by DCE requires making additional conceptual and statistical assumptions on the relationship between detection probabilities and true income. Because the estimates of undetected evasion are large, this issue creates significant uncertainty. In Guyton et al. (2023), which is included in the Online Appendix, we discuss this methodological challenge in depth, and we present results employing an array of different methods for distributional analysis building on DCE. These methods include the methods employed by Johns and Slemrod (2010) and the methods used in IRS (2007) and IRS (2019) to map total under-reporting to the tax gap (which requires some distributional assumptions due to the non-linearity of the tax schedule). The current microsimulation methods in IRS (2019) use an allocation of evasion at the individual level based on imputed mean amounts of under-reporting which, by limiting individual heterogeneity in amounts of under-reporting, will tend to reduce evasion at the very top end of the distribution even if the underlying structural model of DCE is correctly specified.[47] This concern and the broader concern that DCE methods were not designed for analysis focused on the very top of the income distribution are our motivation for considering a variety of methods here.

Methodology to incorporate DCE. We incorporate current DCE methodology while noting its limitations for distributional analysis. We begin by estimating aggregate income under-reporting according to the estimated DCE model, using the methods from the most recent tax gap study, IRS (2019). These estimates pertain to tax years 2008–2013, because IRS (2019) used random audit data for this period to implement DCE. We adapt these estimates to our full sample period of 2006–2013 where possible, assuming that rates of under-reporting for each type of income are similar for tax

[46] "Detectable" here refers to whether detecting some type of non-compliance is feasible given audit procedures and information and resource constraints. The boundary between "detectable" and "undetectable" depends on the design of the audit program.

[47] We thank Brian Erard for helping us to understand this point and the subtleties of DCE generally.

years 2006–2013 and 2008–2013.[48]

We then allocate undetected under-reporting identified by DCE through the income distribution, confronting the challenge described above. To illustrate how uncertainty about the distribution of undetected under-reporting shapes estimates of other distributional quantities of interest, we consider multiple distributional specifications here, holding the total under-reporting constant at the level implied by the aggregate estimates based on IRS (2019). One set of estimates uses the micro-simulated allocation of under-reporting used in IRS (2019) to map DCE-identified under-reporting to the tax gap. Because this specification was employed in the 2019 Tax Gap study, we label this the *DCE2019* specification. We also present results for an allocation method in which we allocate total DCE-identified evasion by type of income according to empirical income shares. We call these methods *Macro Allocated DCE,* or *MA-DCE.* Our preferred, benchmark MA-DCE uses exam-corrected income shares to allocate DCE adjustments. This specification is as close to distributionally neutral as possible, while still accounting for the slight change in the composition of overall income by type of income implied by the aggregate estimates. In sensitivity analysis, we further consider alternative Micro and Macro-Allocated DCE specifications, which we discuss in further detail in Guyton et al. (2023). These include the methods used on 2001 NRP data in IRS (2007) and Johns and Slemrod (2010), which we label *DCE2001*, and several alternative macro allocation specifications based on empirical income distributions (e.g. the distribution of exam-detected under-reported income). In Section 6 below, we also present all our main distributional estimates with and without undetected under-reporting identified by DCE, to illustrate transparently how this component of the estimates shapes the results.

5.2 Results Incorporating DCE-Identified Under-Reporting

Aggregate Estimates. We replicated the analysis of individual under-reporting that generates the estimate of the individual filer tax gap in the most recent IRS Tax Gap study (IRS, 2019).[49] With DCE, we estimate that 10.7% of true income is under-reported, of which 3.8% was detected in exams and 6.9% is undetected evasion identified by DCE methods. In 2012 dollars, we estimate that $975 billion of income was under-reported and $300 billion dollars of individual income tax (including self-employment taxes and refundable credits) was unpaid on average over 2008–2013. The latter figure closely matches the official tax gap figure for 2011-2013 in IRS (2019), $290 billion.[50]

Appendix Table A6 and Figure A13 describe the aggregate DCE adjustments by type of income. About 80% of the aggregate DCE adjustment comes from forms of income that are moderately

[48] In some sensitivity analysis, we compare different specifications for 2008–2013 rather than 2006–2013 to ensure that estimates are comparable across specifications; unless otherwise noted, estimates pertain to 2006–2013. Which years one uses has a level effect when we turn to the income inequality estimates in Section 6 but our main findings are robust.

[49] We thank Drew Johns for sharing the relevant simulations and for helping us to understand the underlying methods.

[50] See (IRS, 2019, Table 2 p. 11). The individual income underreporting tax gap is $245 billion and the self-employment under-reporting tax gap an additional $45 billion.

or highly concentrated at the top of the distribution: Schedule C income, capital gains, rental income, pass-through business income, and line 21 other income (e.g. net operating loss carry-forwards).[51] Less than 10% of the aggregate adjustment comes from wage and pension income, which are relatively more equally distributed.[52] As a result, we observe, comparing Table A1 to Table A6, that DCE adjustments tilt the composition of under-reported and true income toward forms of income that are concentrated at the top of the distribution.

Distributional Methods. The key unknown in estimating the location in the income distribution of undetected evasion identified by DCE is the relationship between the probability of detection and true income. Pinning down this relationship requires more assumptions than the necessary assumptions to identify total under-reporting described above. We approach this challenge in two different ways, implementing one approach using micro simulations (Micro DCE) and a more macro approach (MA-DCE) based on other empirical income distributions. We view these two types of approaches as complements. Adopting a micro simulation approach makes micro-level information like the influence of re-ranking on the estimated profile of evasion transparent, but it can be difficult to assess what structure the model imposes on the crucial relationship between detection probabilities and true income. With a macro approach, the opposite is true: one makes transparent assumptions governing the relationship between detection and true income, but what the model implies about re-ranking becomes opaque.

The micro-simulation method used in IRS (2019) assigns amounts of undetected under-reporting to each individual tax return; precisely how they do so imposes structure on the relationship between detection probabilities and and true income. Researchers first estimate a structural model of DCE using a model based on Erard and Feinstein (2011). Based on the estimated model and the contents of each individual tax return (as originally filed), every income line item on each individual tax return is allocated 1) a probability of containing under-reported income, and 2) an amount of expected under-reporting that will be allocated to them conditional on under-reporting. The simulation then takes random draws to specify whether each taxpayer under-reported each type of income given their assigned probability of under-reporting.[53] To obtain an estimated profile of under-reporting, we rank individuals in the simulated data by their rank in the true income distribution, and then estimate distributional statistics.

Note that individuals simulated to be under-reporters were allocated a specific dollar amount of under-reporting representing a conditional expectation, *not* a random draw from a conditional distribution. This feature of the model implies that there will be too few large-dollar amounts of

[51] In 2012, the top 1% of tax filers earned 22.7% of reported market income, 19% of schedule C income, 83% of capital gains, 78% of pass-through business income, and more than 100% of rental income (which on aggregate was negative).

[52] In 2012, the top 1% of tax filers earned 12% of wage income and 6% of taxable pension income excluding Social Security.

[53] As described in IRS (2019), the simulations were run 10 times; following the same approach, we take an average across all 10 simulations in estimating aggregate statistics.

evasion in the simulation, which implies that too little evasion will wind up at the top of the distribution after re-ranking, even if the under-lying structural model in DCE2019 is correctly specified.[54]

In addition to this approach, we illustrate in sensitivity analysis a second micro approach based on DCE methods circa the 2001 NRP, which we call DCE2001 (Johns and Slemrod, 2010). With DCE2001, one first estimates the ratio of total under-reporting to detected under-reporting within classes of types of income and (originally filed) tax returns using the auditor effects model. One then scales detected under-reporting *at the micro level* by these ratios, or multipliers. We discuss further how the underlying methods in each of these approaches imposes structure on the relationship between detection probabilities and true income further in Guyton et al. (2023). The approach allocates more undetected evasion to individuals with large amounts of detected evasion, which leads to substantially more re-ranking. Because there may be some individuals with large amounts of detectable but undetected evasion and little to no detected evasion, we should regard skeptically the re-ranking that occurs with DCE2001; however the direction of the bias implied by this approach is *ambiguous* (see Reck et al., 2021, for a demonstration of the ambiguity).

Because of the concerns with either micro approach, we also implement what we call a macro allocation DCE (MA-DCE) approach. With this approach, we start from the exam-corrected data from Section 2.2 and we use empirical moments from the distribution of other types of income to discipline the allocation of the additional under-reported income identified by the DCE model. In our baseline MA-DCE specification, we assume that for each type of income, the under-reported income identified by DCE is distributed like exam-corrected income for that type of income. For example, if X% of interest income belongs to the top 1% of the exam-corrected income distribution according to individual random audit estimates from Section 2.2, we allocate X% of DCE-identified undetected under-reported interest income belongs to the top 1% of the true income distribution. Note that with this allocation, X% of interest income belongs to the top 1% *before and after* we add MA-DCE adjustments to exam-corrected estimates, so this method is distributionally neutral for each type of income by design.[55] In sensitivity analysis, we replace the exam-corrected income shares with other income shares reflecting alternative possibilities. We thereby show how it would impact the estimates to suppose that the undetected under-reporting identified by DCE is distributed just like reported income or exam-detected under-reported income. We do all of these separately by type of income, but we also consider a specification in which we pool some line

[54] The primitives of the underlying structural model are posed in terms of log income. The use of an anti-log transform to obtain estimates denominated in dollars creates some additional concerns, for example compressing the distribution of undetected misreporting toward the population mean, which we discuss further in Guyton et al. (2023).

[55] We view this as a desirable property given the absence of direct evidence about the location of the undetected under-reporting identified by DCE models in the true income distribution. We will see below that estimates of corrected top fiscal income shares depend on how this evasion is distributed. Distributing DCE adjustments in an approximately neutral fashion limits the effect of these adjustments on the overall estimated distribution of income, e.g. on the top 1% income share. As we discuss in Guyton et al. (2023), we do not believe there is compelling evidence to support a substantial revision of the top 1% share in either direction due to the evasion identified by DCE.

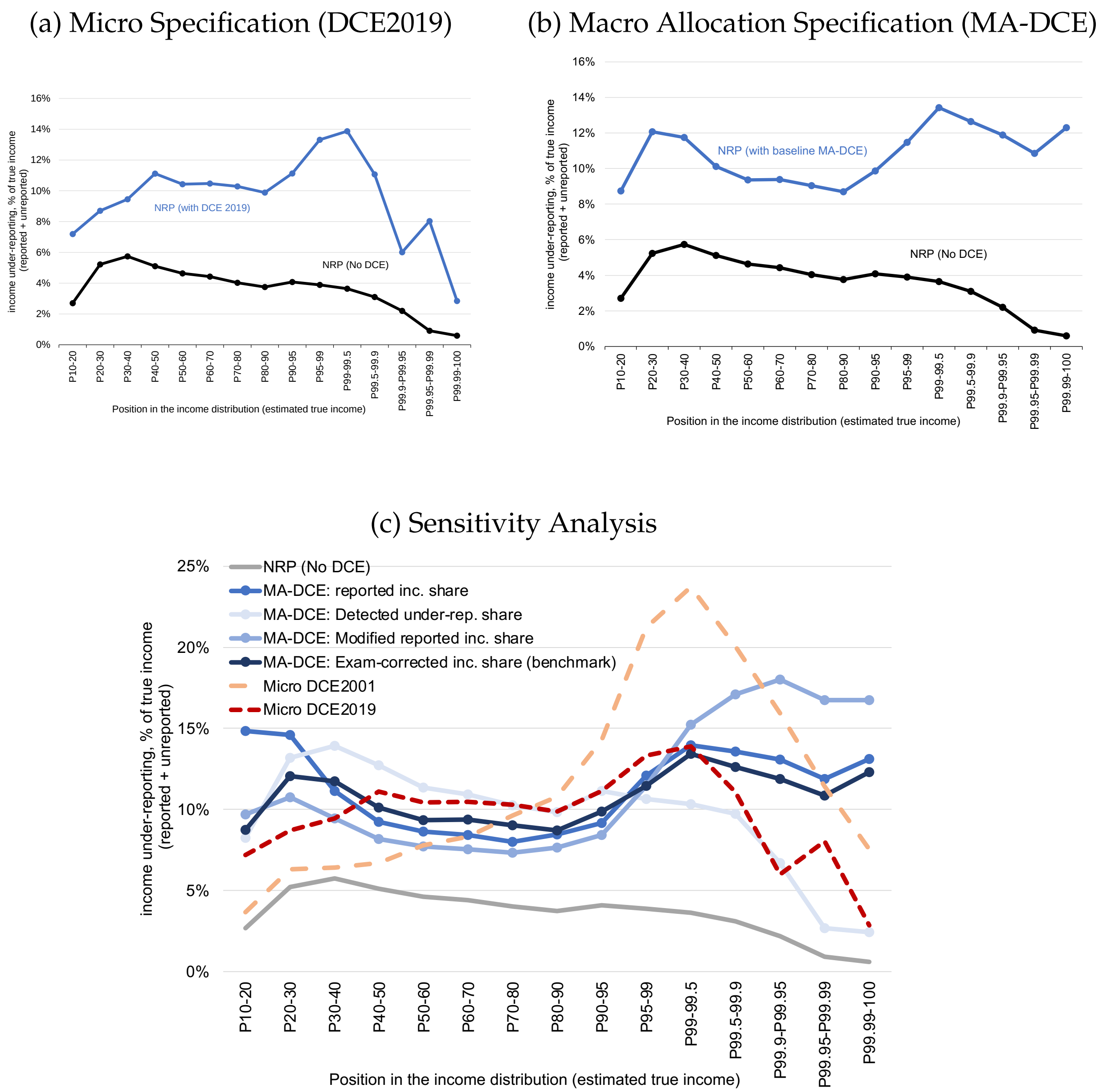
items of income and regard them as a single category of income. Considering a range of macro allocation specifications facilitates comparisons to micro specifications.

Distributional Estimates. Figure 8 illustrates the estimated profile of evasion by rank in the true income distribution, including DCE-identified under-reporting (but without including additional under-reporting based on Sections 3 and 4).

Figure illustrates the profile of evasion using 2019DCE, and Figure 8b illustrates the baseline MA-DCE method based on exam-corrected income shares. With 2019DCE, under-reporting increases from about 7% of true income at the bottom to about 10% of true income from the 30th to the 90th percentile of the income distribution. It increases further to about 13% of income at around the 99th percentile, and then decreases sharply within the top 1% of the distribution. The sharp dropoff occurs because 1) exam-detected under-reporting is rare at the top (see Figure 1a) and 2) this method assigns "typical" amounts of exam-detected under-reporting to simulated individual under-reporters. With the more distributionally neutral MA-DCE specification, about 13% of true income is under-reported at the bottom of the distribution. This rate falls to about 8% for most of the middle of the distribution and then increases to about 14% in the top 1%. Because overall DCE-identified under-reporting consists of relatively concentrated types of income, the difference between exam-detected under-reporting and estimated overall under-reporting is largest at the top of the distribution with this specification.

Figure 8c reports the results of a more detailed sensitivity analysis, wherein we make different assumptions about the distribution of DCE-identified under-reporting. We illustrate the two specifications from Figures 8a and 8b alongside DCE2001 and several alternative macro allocation specifications. For the macro allocation specifications, we obtain similar results to our baseline MA-DCE method if we distribute DCE-identified under-reporting like reported incomes rather than exam-corrected incomes. We also consider a specification in which we assume that DCE-identified under-reporting is distributed like exam-detected under-reporting. Compared to other specifications, under-reporting is smaller at the top of the distribution and falling within the top 1% with this specification, which reflects that under-reporting detected during individual random audits is also much lower at the top of the distribution (recall Figure 1a). As discussed further in Guyton et al. (2023), we regard this scenario as a lower bound with respect to the concentration of under-reporting because 1) the character of income and third-party reporting through the distribution suggests that the probability of detection is likely flat or decreasing through the distribution (see Figure 1b), and 2) incorporating additional under-reporting by those with some under-reporting in exam-corrected data should cause re-ranking effects that move under-reporting upward in the distribution, effects which are not captured by this specification. We also report the results from a high-end scenario in Figure 8c, wherein we assume non-compliance is similar in sole proprietorships, partnerships, and S corporations, and thus regard these as a single type of income when distributing aggregate DCE-identified under-reporting by type of income.

Figure 8: Detection-Controlled Estimation: Estimated Under-Reported Income Including DCE-Identified Evasion



Note: Panel a) of this figure illustrates how incorporating DCE-identified under-reporting of income using the DCE2019 micro simulations affects the profile of under-reporting. Panel b) does the same thing using our baseline MA-DCE method – allocating DCE-identified under-reporting like exam-corrected incomes by type of income. Panel c) plots the results from several alternative specifications, to illustrate how varying assumptions about the distribution of undetected under-reporting shape estimates of the overall rate of under-reporting of income. We include marco allocation estimates based on assumptions that DCE-identified under-reporting is distributed like some other type of income (e.g. reported income), and micro estimates that allocate DCE-identified under-reporting to individual taxpayers, using two specifications used in prior tax gap studies. We label the macro allocation approaches according to the type of income used to allocate DCE-identified evasion, and we label the micro approaches DCE2001 and DCE2019 as discussed in the main text. See Guyton et al. (2023) for further details. We contrast all of these with the profile of under-reporting estimated based on exam corrections only, without including any undetected under-reporting (c.f. Figure 1a).

One lesson we take away from this sensitivity analysis is that the underlying simulations in DCE2019 implicitly impose that DCE-identified undetected under-reporting is distributed in a similar fashion to exam-detected under-reporting. When we implement a macro allocation specification that distributes DCE-identified under-reporting according to the share of exam-detected under-reporting that belongs to a given part of the income distribution, we arrive at a profile of evasion that resembles that of DCE2019, especially at the top of the distribution. In Guyton et al. (2023), we conceptually connect this similarity to the underlying simulation method. Essentially, the DCE2019 simulations are using exam-corrected data to identify what is a typical exam correction for a tax return with a given set of characteristics, and then simulating a model in which many individuals – more than were actually detected, due to auditor effects – under-report by these typical amounts. The use of simulated distributions for the extensive margin of evasion (whether someone under-reports) but not the intensive margin (how much someone under-reports conditional on under-reporting) will lead to too few large amounts of undetected under-reporting and overly modest re-ranking effects.

We also include in Figure 8c the specification used on the 2001 NRP data in IRS (2007) and Johns and Slemrod (2010), i.e. DCE2001. This specification yields a a similar profile of evasion to our baseline MA-DCE method when we compare the top 1% to the rest of the population, but the DCE2001 specification features a steeper rise in evasion from the 90th to 99th percentile of the distribution, and a steep decline in evasion at the very top. Our results with this specification closely resemble those of Johns and Slemrod (2010). In Guyton et al. (2023), we trace the steep increase and decline in the rate of under-reporting to the re-ranking effects induced by the multiplier method of DCE2001. Reck et al. (2021) provides additional sensitivity analysis using similar micro methods to DCE2001.

In summary, estimated rates of under-reporting are sensitive to alternate assumptions about the distribution of the undetected under-reporting identified by DCE, particularly within the top half of the top 1%. Holding aggregate under-reporting fixed, the fact that the very top of the distribution comprises far fewer individuals than the bottom 99% generates more quantitative uncertainty about the rate of under-reporting at the top. Specifications that differ by one or two percentage points in the rate of under-reporting in the bottom 99% can differ by perhaps five percentage points in the top 1%.

6 Sophisticated Evasion and the United States Income Distribution

In this section, we present new estimates of the distribution of noncompliance in the United States. We augment the estimates used by the IRS in its tax gap studies (e.g. IRS, 2019), as summarized in Section 5, to account for the forms of evasion studied in Sections 3 and 4, forms which are virtually never detected during individual random audits given the audit procedures, the available information, and resource constraints. We provide two benchmark scenarios – using either DCE2019 or the baseline MA-DCE method as the starting point for incorporating offshore and pass-through evasion – along with extensive sensitivity analysis. The three main findings are the following. First, while our benchmark estimates feature only slightly more evasion on aggregate than in official statistics, our proposed adjustments have large effects at the top of the income distribution. Second, in our benchmark estimates, we find that rates of income under-reporting rise with true income up to the 95^{th} percentile of the distribution, and above this level under-reporting is high and either roughly constant or modestly declining as a share of true income. Third, we find that accounting for unreported income increases the top 1% income share for 2006–2013 significantly, by around one percentage point.

6.1 Methodology

Throughout, we classify evasion in two categories: sophisticated and less sophisticated. In our benchmark scenarios, we estimate less sophisticated evasion using the results from Section 5, which are based on the official tax gap estimates of overall income under-reporting, plus additional distributional assumptions of either DCE2019 or our baseline MA-DCE method. That is, less sophisticated evasion comprises estimated noncompliance detected in NRP individual random audits, adjusted to account for differences in experience (and other observable characteristics) across examiners (see Figure 8). To understand how some of the assumptions around DCE discussed in the previous section matter, we sometimes split less sophisticated evasion into an exam-detected component and a component an an undetected component due to DCE. Next, we incorporate sophisticated evasion as the sum of offshore and pass-through evasion, using our benchmark scenario for each (Figures 4a and 7a). We combine sophisticated and less sophisticated evasion additively at the distributional level for simplicity, i.e. assuming no re-ranking when adding them. While doing so, we account for the potential overlap between the DCE adjustment and our proposed adjustment for sophisticated evasion. Specifically, we remove 57.6% of the post-DCE estimate of partnership and S corporation evasion from the NRP data, before adding our estimate of entity-level pass-through evasion.[56] In both benchmark scenarios, total under-reported income is 12.1% of total true income: 2.0% from sophisticated evasion and 10.1% from less sophisticated evasion.[57]

This benchmark scenario is motivated by the following considerations. First, random audit data are the best tool available to measure less sophisticated forms of evasion. Moreover, the fact

[56] As we have seen (footnote 41), 57.6% of partnership and S corporation income evasion detected in the NRP is associated with an entity pick up (i.e., an audit of the corresponding business). Therefore, up to 57.6% of DCE-adjusted pass-through income evasion in the NRP can be seen as capturing business-level evasion. We remove all this business-level evasion before adding our own estimate of pass-through-business-level evasion.

[57] Total under-reported income (as a fraction of true income) for less sophisticated evasion (10.2%) is slightly lower than total DCE-adjusted NRP evasion (10.7%) because (i) we remove 57.6% of pass-through evasion; (ii) the "true income" denominator is enlarged by adding sophisticated evasion. Similarly, total under-reported income for sophisticated evasion (2.0%) is slightly lower than total offshore plus pass-through evasion (2.1% as a fraction of exam-corrected, non-DCE adjusted income) because the denominator is enlarged by the DCE adjustment.

that more experienced auditors systematically detect more evasion suggests that some evasion is missing in the raw exam-corrected data. The estimates incorporating DCE are the best estimates currently available to capture the insight that examiners vary in their capacity to uncover noncompliance; thus the estimates from the previous section are a natural starting point for less sophisticated evasion. Last, as we have seen, due to the practical limits inherent to the procedures and constraints of random audits during our sample period, these estimates largely misses offshore and pass-through business evasion. This is true even of the estimates including DCE adjustments, because detected noncompliance of offshore or entity-level pass-through evasion virtually never occurs during the underlying audits.

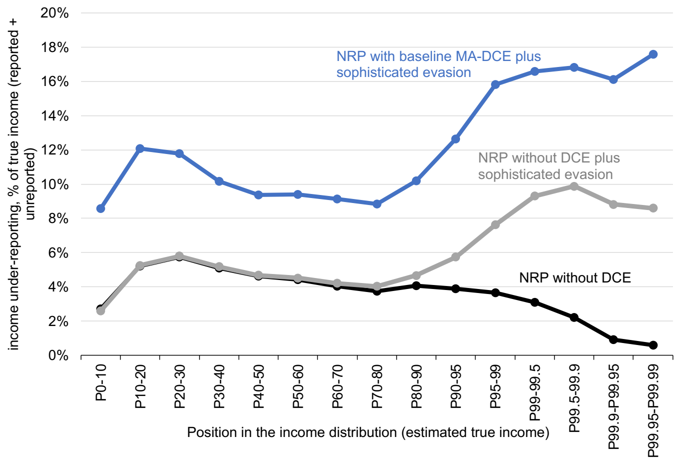
Because measuring evasion—especially undetected evasion—necessarily involves a margin of error, we also consider a large number of sensitivity tests capturing the key dimensions of uncertainty. One set of sensitivity tests concerns uncertainty about the size of sophisticated evasion. First, we implement the high-end and low-end scenarios for offshore and pass-through evasion described in Figures 4b and Figure 7b, respectively. Second, we consider the case where sophisticated evasion on aggregate is lower or higher than 2.0% of true income but distributed like in our benchmark series. For the lower case, we assume that sophisticated evasion adds up to 1.5% of true income. This reflects a post-FATCA world in which FATCA would be fully effective (so that under-reporting rates on offshore capital income would be reduced by a factor of 4 relative to our benchmark scenario and become similar to the under-reporting rates on onshore capital income) while other forms of sophisticated evasion would remain constant. For the higher case, we consider a scenario in which total sophisticated evasion adds up to 2.5% of true income. This scenario reflects the possibility that there are quantitatively significant forms of sophisticated evasion other than pass-through and offshore evasion, such as abusive uses of trusts, charities, and tax shelters used by high-net-worth individuals. Another set of sensitivity tests concerns uncertainty about undetected evasion accounted for by DCE methodology. To transparently show how DCE affects the results, we show a version of all our results without DCE adjustment. We also illustrate our main results for each of the alternative assumptions about the location of DCE adjustment through the income distribution, as in Figure 8c. As we shall see, there is quantitative uncertainty around rates of under-reporting at the top, but all our main findings are robust to these variations.

6.2 Main Results

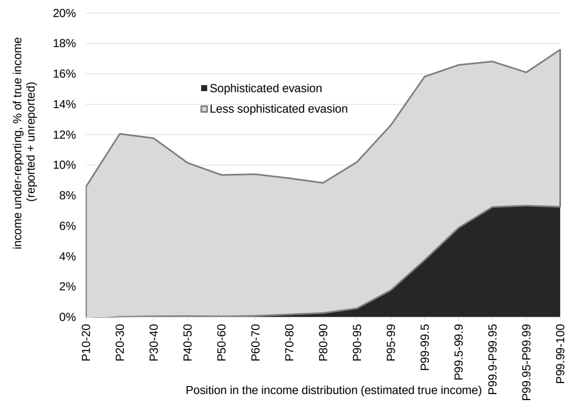
Figure 9 depicts estimated income under-reporting in our benchmark scenarios. We find that income under-reporting hovers around 10% of true income in the bottom 80% of the distribution, rises to around 16% of true income from the 80^{th} to the 99^{th} percentiles, and then remains fairly constant or modestly declines within the top 1 percent, depending on which DCE specification we use. Figures 9b and 9d decompose these estimates into sophisticated versus less sophisticated evasion. Sophisticated evasion is negligible in the bottom 90% of the distribution, but it rises to over 7% of true income in the top 0.01%.

FIGURE 9: THE DISTRIBUTION OF NONCOMPLIANCE IN THE U.S.: BENCHMARK ESTIMATES

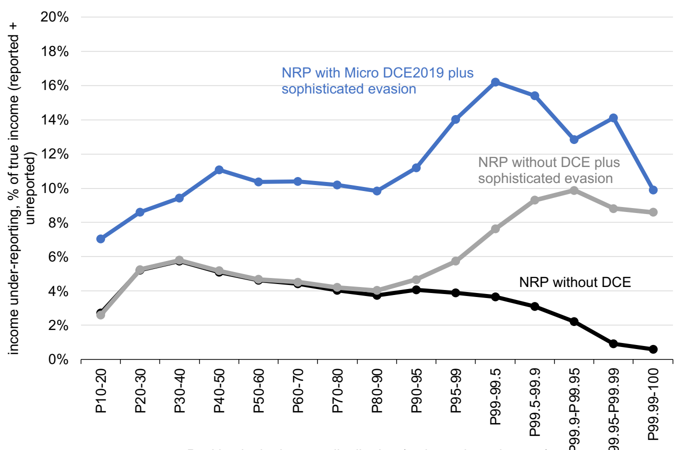
(a) Unreported Income (% True Income) Using Macro Allocated DCE



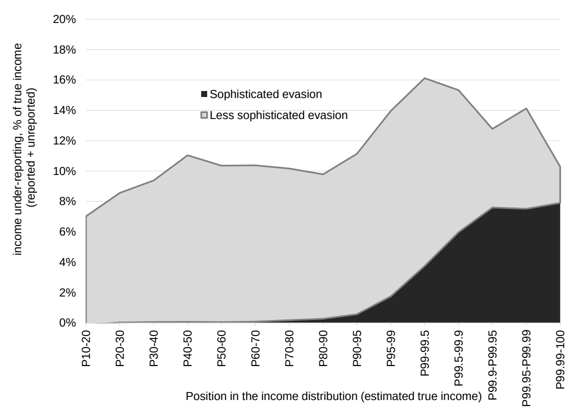
(b) Composition of Unreported Income Using Macro Allocated DCE



(c) Unreported Income (% True Income) Using Micro DCE



(d) Composition of Unreported Income Using Micro DCE



Note: This figure plots estimates of under-reported income by true income groups, combining evasion detected in NRP random audit data, evasion identified by DCE methods, and sophisticated evasion (offshore and pass-through evasion) under our our benchmark assumptions for tax year 2007. The top two panels employs our baseline MA-DCE and the bottom two panels employ DCE2019. For each of these, the first panel illustrates the components of the benchmark estimate. Beginning with evasion detected by random audits (before DCE adjustment as in Figure 1a), we add sophisticated evasion (combining the results in Figures 4 & 7a), and then we incorporate evasion identified by DCE (building on Figure 8b). We rank individuals by estimated true income either before or after DCE adjustment. The second panel decomposes the benchmark estimate into its sophisticated component, which includes offshore and pass-through evasion, and the residual, "less sophisticated" component. For details on each specific components of the benchmark estimates, see the notes to previous figures.

On aggregate, our benchmark income under-reporting gap is only about 1.19 times larger than the gap estimated after DCE in IRS (2019). However, our benchmark adjustment increases the income under-reporting gap by a factor of about 1.5 for the top 1%. This is because sophisticated evasion, although not large on aggregate in our benchmark estimates, is highly concentrated, with 75% of it going to the top 1%. Second, building on the results in Sections 3 and 4, in Figure 9 we also show how sophisticated evasion modifies the profile of evasion detected in NRP random audits, without incorporating undetected evasion identified by DCE. Without DCE, the effects of adding sophisticated evasion are even more stark. While in non-adjusted random audit data, the highest earners under-report much less income (relative to true income) than other taxpayers, after adding sophisticated evasion the opposite is true. Under-reported income rises from 4%–6% of true income in the bottom 95% of the income distribution to 8%–10% in the top 1%.

Table 1 reports the share of total unreported income attributable to various income groups. We estimate these quantities starting from NRP data for tax years 2008-2013 because these are the years for which we are able to implement DCE2019 using the methods from IRS (2019). Appendix Table A7 reports results based on our full sample period of 2006–2013, omitting DCE2019. We estimate that unreported income is very slightly more concentrated across 2008–2013 than across 2006–2013 across all specifications, but the rest of our findings are virtually identical.

Altogether, we estimate that 24.4% of unreported income is earned by the top 1% using micro DCE2019; with our baseline MA-DCE specification this figure is 31.9%. When disregarding sophisticated evasion (i.e., in the DCE-adjusted random audit data), unreported income is slightly less concentrated: the top 1% earns 15.3% of unreported on average over 2008–2013 under the micro DCE2019 specification and 24.9% with the baseline MA-DCE specification. This compares to a 27% estimate for the 2001 NRP data in Johns and Slemrod (2010).[58] We also report the distribution of unreported income when disregarding DCE adjustments. In that case, without sophisticated evasion, the top 1% earns 13.0% of all under-reported income. Adding sophisticated evasion, the top 1% earns 34.9% of all unreported income. In summary, adding sophisticated evasion increases the estimated share of unreported income earned by the top 1% by 7 percentage points under the baseline MA-DCE specification, by 9.1 pp under the DCE2019 micro specification, and by 21.9 pp when disregarding DCE-identified under-reporting.

Our findings on the income under-reporting gap naturally carry over to the tax gap. When accounting for sophisticated evasion, the tax gap is higher than officially estimated in IRS (2019), by a factor of 1.15.[59] As sophisticated evasion is concentrated at the top, more than 80% of this increase is in the top 1%.[60]

[58] Using the older DCE allocation method that Johns and Slemrod used on the on the 2001 NRP, we estimate that the top 1% earns 25% of all under-reporting (not accounting for sophisticated evasion), see Guyton et al. (2023).

[59] In our data, the annual tax gap for the federal income tax (excluding refundable tax credits but including self-employment taxes) is $259 billion on average over 2008-2013 (in constant 2012 dollars), before accounting for sophisticated evasion. This number is consistent with the one reported in the official tax gap publication (IRS, 2019) for 2011-13, $248 billion. Accounting for sophisticated evasion adds $39 billion to this total according to our estimates.

[60] Uncertainty around the allocation of DCE adjustments complicates the mapping of income under-reporting at the top to the tax gap at the top (see Guyton et al., 2023). Without DCE adjustments, accounting for sophisticated evasion increases the share of unpaid tax attributable to the top 1% from 13% to 33%. When further incorporating DCE adjustments in various ways, this share remains in the neighborhood of 33%. The top 1% taxpayers remitted about 35% of all

6.3 Implications for the Measurement of Inequality

There has been renewed interest in recent decades in the study of income and wealth distributions. Income tax returns are a key data source for the study of inequality. However, the tax data used to study inequality in the United States are before any correction for unreported income. How does accounting for unreported income affect what we know about inequality?

Table 2 reports the distribution of taxable market income on average over 2008–2013 with different treatments of unreported income. Table A8 reports analogous estimates for 2006–2013 where possible, as above. Inequality is slightly lower in 2008–2013 because the years of the financial crisis get more weight with this range of years; apart from this level difference we obtain the same results for either period.

The first column shows the distribution of reported income with no adjustment for evasion. The next two columns shows the distribution of reported plus detected under-reported income (i.e., with no adjustment for undetected evasion). This distribution is similar to the distribution of reported income, because only 4.0% of income is found to be under-reported on aggregate in random audit data. Because under-reporting is less concentrated at the top than reported income, accounting for evasion modestly decreases the top 1% share, by about 0.4 percentage points. The next two columns incorporate adjustments for DCE-identified evasion.[61] After DCE adjustment, we estimate that the top 1% income share falls slightly with micro DCE2019 specification. With the baseline MA-DCE specification, the top 1% income share instead rises slightly, because the forms of income receiving large DCE adjustments are more concentrated at the top than overall income. The fifth column adds sophisticated evasion to random audit estimates without DCE, which increases the top 1% income share from 19.1% to 19.9%. The last two columns of Table 2 show the distribution of true income including both DCE-identified and sophisticated evasion. If we adopt the DCE2019 micro specification, accounting for sophisticated evasion increases the top 1% income share from 18.1% to 19.2%. If we adopt the baseline MA-DCE specification, accounting for sophisticated evasion increases the top 1% income share from 19.5% to 20.4%. In any case, after we account for sophisticated evasion, the top 1% income share is similar to or up to 1.2 pp higher than one would estimate based on reported incomes alone.

These results have implications for estimating the distribution of total U.S. national income, as in Piketty et al. (2018) and Auten and Splinter (2019). For the computation of U.S. national income, the Bureau of Economic Analysis includes income which should be reported in tax returns but is not. For wages, sole proprietorship income, and partnership business income, the amounts of

federal income taxes in this period, so overall our results suggest that the top 1%'s share of unpaid tax is similar to their share of paid taxes.

[61] These columns are most directly comparable to the results in Johns and Slemrod (2010) on how accounting for evasion modifies the estimated income distribution. For the 2001 NRP, Johns and Slemrod (2010) found that accounting for evasion changed the top 1% share by less than 0.1 percentage point. See Appendix Tables A9 and A10 for side-by-side comparisons, and refer to Section 5 and Guyton et al. (2023) for further discussion.

unreported income included in national income are directly based on the post-DCE NRP.[62] For other income categories, the inclusion of unreported income is implicit; for instance, aggregate rental income is higher in the National Income and Product Accounts (NIPAs) than in tax returns in part because some rental income is unreported in individual income tax returns.[63] Since the top 1% income share is higher after allocating unreported income, our results in this paper suggest that accounting for tax evasion should increase income concentration in 2006–2013. In Auten and Splinter (2019), the top 1% income share is lower by 0.8 percentage points after the allocation of unreported income on average over 2006–2013 (and by 0.4 percentage points in 2001).[64] In Piketty et al. (2018), the top 1% income share is higher by 0.7 percentage point after including the forms of evasion explicitly identified in the NIPAs, with no time trend.

6.4 Sensitivity Analysis

We next conduct sensitivity analysis for all the key results from this section. As discussed above, one set of sensitivity checks pertains to sophisticated evasion and the other set pertains to the allocation of evasion identified by DCE methodology.

We present the results of the sensitivity analysis around the location in the allocation of DCE-identified evasion in Appendix Figure A14 and Tables A11 through A14. As above, we include results that start from NRP data from 2006–2013, our full sample period, and from 2008–2013, where we can implement DCE2019. In the lower-bound based on the detected under-reporting share macro allocation DCE specification, less sophisticated evasion in the top 1% is relatively low, from about 8% of true income, and incorporating sophisticated evasion increases this to about 12%. In the high-end scenario based on the modified reported income share MA-DCE specification, accounting for sophisticated evasion increases the rate of under-reporting at the top 1% from 16% to 22% of true income. In other words, uncertainty about the distribution of DCE-identified evasion accounts for 8-10 percentage points worth of uncertainty in either direction for the rate of under-reporting at the top, but our main finding – accounting for sophisticated evasion substantially increases the top 1% under-reporting rate – obtains for any reasonable DCE scenario. Likewise uncertainty about DCE implies some uncertainty about the concentration of unreported and true income, but it does not reverse our findings that accounting for sophisticated makes an

[62]In 2013, the adjustment amounts to $76.8 billion for wages (NIPA Table 7.18 line 2) and $639.8 billion for sole proprietorship and partnership income (NIPA Table 7.14 line 2).

[63]No explicit reconciliation between aggregate rental income in the NIPAs and in 1040s is provided by the Bureau of Economic Analysis because NIPA rental income is not estimated based on tax returns. Other conceptual differences (e.g., the treatment of depreciation) contribute to the gap between NIPA income and income reported in tax returns; see Saez and Zucman (2020, section 3.1.1).

[64]For example, in Auten and Splinter (2019), the top 1% income share before the allocation of under-reported income is 16.9% in 2013 (Table C1-Incomes, col. DE divided by col. DB) and 16.1% after adding unreported income (Table C1-Incomes, col. DM divided by col. DJ). Out of this 0.8 percentage point decline, about 0.6 percentage points comes from tax filers. An additional 0.2 percentage points come from non-filers, as Auten and Splinter (2019) allocate 15% of their total under-reported income aggregate to non-filers, assuming this income is earned by people at the bottom of the true income distribution. We do not analyze non-filer income in our paper, but we note that recent analysis shows that high-income non-filers drive much of the non-filer tax gap in recent years (TIGTA, 2020).

important difference, nor that accounting for evasion increases the concentration of income relative to reported incomes. The estimated top 1% share of unreported income ranges from 24% to 41% for 2008–2013, compared to the unadjusted NRP estimate of 13%. The estimated top 1% income share of true income ranges from 19.2% to 21.5% for 2008–2013, compared to the estimate based on reported incomes alone of 19.1% and an estimate based on exam-corrections of 18.7%.

We present the results of sensitivity analysis around sophisticated evasion in Appendix Figure A15 and Tables A15 through A20. The basic pattern of under-reporting depicted in Figure 9b remains across all scenarios we consider. Under-reporting rises with income up to about the 99^{th} percentile of the income distribution, where it reaches 15% of true income or slightly more. In our most extreme scenario, using the upper bound for sophisticated evasion and MA-DCE with modified reported income shares, the evasion rate rises sharply within the top 1%; the top 0.01% under-reports about 25% of its true income. In the least extreme scenario, using the lower bound for sophisticated evasion and micro DCE2019, evasion is declining with income, and the top 0.01% evades about 6% of its income. Across all of these, sophisticated evasion constitutes an important modification to the concentration of under-reported income, and the concentration of of true income. The estimated top 1% share of unreported income in 2008–2013 ranges from 18% to 33% under micro DCE2019 and from 27% to 39% under baseline MA-DCE, compared to the unadjusted NRP estimate of 12%. The estimated top 1% share of true income in 2008–2013 ranges from 18.4% to 20.4% under micro DCE2019 and from 19.7% to 21.7% under baseline MA-DCE, compared to an estimate based on reported incomes alone of 19.1%, and an estimate based on exam-corrections of 18.7%.

7 Theory

In this section, we inform the interpretation of the above body of empirical results with some simple economic theory. Our central goal is to explain why some forms of sophisticated evasion are concentrated at the top of the income distribution. We also briefly consider some related optimal tax enforcement questions.

7.1 Model of Endogeneous Concealment

Setup. As in Allingham and Sandmo (1972), an individual determines how much income to report to the tax authority out of exogenous true income, y. Evaded income is denoted by e. The only modification is that we give the agent the option to take a binary concealment action $a \in \{0, 1\}$ that will reduce the probability of detection of evasion at some fixed cost κ. The individual's optimization problem is:

$$\max_{e \in [0,y], a \in \{0,1\}} (1 - p(a))u((1 - \tau)y + \tau e - \kappa a) + p(a)u((1 - \tau)y - \tau\theta e - \kappa a) \tag{1}$$

where τ is the tax rate, $\theta > 0$ is the penalty rate (following Yitzhaki, 1974), $u()$ is standard risk-averse utility over ex post consumption with $u' > 0, u'' < 0$, and $p(a)$ is the probability of detection with $p(1) < p(0)$. If we restrict the individual to $a = 0$, this model obviously becomes the Allingham-Sandmo model.

We view this as an intuition-building model, so there are multiple interpretations of the action a. For example, the concealment action could be shifting income offshore or exploiting a complex pass-through structure to tax obligations. With this interpretation, one can think of $p(a)$ as the probability of audit times the probability that the auditor discovers the taxpayers' evasion conditional on a, $p(a) = \Pr(\text{audit}) * \Pr(\text{auditor detects evasion}|a)$. Prior literature implicitly assumed the second probability in this expression is one, even when adopting an endogenous detection probability (Kleven et al., 2011). We could also interpret a as the adoption of a gray area avoidance/evasion position that is uncertain to be successful if the position is legally challenged. In this case, the model looks exactly the same, but we would impose $p(a) = \Pr(\text{audit}) \Pr(\text{taxpayer loses legal challenge}|a)$.[65]

For simplicity, we do not directly model third-party information, though it has been shown to be important and straightforward to incorporate into the Allingham-Sandmo framework (Kleven et al., 2011). Intuitively, when information in a taxpayer's income is acquired by the tax authority from a third party, evasion on this income becomes easily detectable, and not reporting such income would have a very high probability of detection. One can essentially think of third-party information as imposing an upper bound on evasion, which would be a straightforward extension of the model here. One can also think of the shifting of income away from what would be covered by third-party information (e.g. shifting assets offshore to avoid third-party reporting on domestic financial capital income, before FATCA) as a concealment action.

Finally, we introduce some notation to facilitate exposition of the results. First, we denote the fraction of true income evaded—the analogue to the rate of under-reporting of income in the model—by $g = e/y$. To distinguish the chosen g in the optimization problem from Equation (1) from the g chosen under $a = 1$ or $a = 0$, we let $g_a(y, p)$ denote the level of g the taxpayer would choose if we maximize the objective restricting to $a = 0$ or $a = 1$, given true income y and probability of detection p. Second, Allingham and Sandmo showed that the effect of changes in income on evasion depend on absolute and relative risk aversion, which we denote by $A(c) \equiv -u''(c)/u'(c)$ and $R(c) \equiv -cu''(c)/u'(c)$, respectively.

[65] In the case of grey area avoidance/evasion, we expect the penalty rate θ to be relatively low, as large penalties are rarely imposed in situations involving legal challenges. Holding all else fixed, decreasing the penalty rate when the individual adopts ($a = 1$) would tend to increase the adoption of grey area avoidance strategies among high-income individuals. More generally, allowing the penalty to depend on θ would not effect the main results here, but a higher penalty conditional on $a = 1$ could negate Assumption 1, i.e. with a high enough penalty, a very high-income taxpayer might have no motive to evade. Our empirical results and the fact that harsh penalties like prison time are very rarely imposed suggest that the case we focus on, where penalties do not completely deter sophisticated concealment, is the empirically relevant one under current policy.

Concealment at High Incomes. We begin by developing some intuition for how the tax gap should vary by income, and how this depends on whether the individual takes the concealment action. The following Lemma summarizes what we know about this from Allingham and Sandmo's analysis of their model, i.e. under $a = 0$:

Lemma 1. ***The Allingham-Sandmo Tax Gap.***

L1.1. *If the individual is risk averse, g_0 is decreasing over p, $\partial g_0/\partial p < 0$.*
L1.2. *If absolute risk aversion is decreasing ($A' < 0$), e is increasing in true income y.*
L1.3. *If relative risk aversion is constant ($R' = 0$), g_0 is constant over true income, $\partial g_0/\partial y = 0$.*
L1.4. *If relative risk aversion is decreasing ($R' < 0$), g_0 is increasing over true income, $\partial g_0/\partial y > 0$.*
L1.5. *If relative risk aversion is increasing, ($R' > 0$), g_0 is decreasing over true income, $\partial g_0/\partial y < 0$.*
Proof. See Allingham and Sandmo (1972). □

Our goal is to capture the intuition that, as income grows very large, the fixed cost of adoption κ becomes a trivial share of income, so the taxpayer will opt for the lower detection probability given the trivial cost. The following assumption ensures that as the cost becomes a trivial share of income, the benefits of adoption do not also become trivial:
Assumption 1. *As y becomes arbitrarily large, $g_0(y, p(1))$ approaches a strictly positive constant.*

Assumption 1 is stated as an assumption about optimal behavior, but it imposes restrictions on the primitives of the model, especially risk preferences. From Lemma 1, we know that Assumption 1 is satisfied under constant and decreasing relative risk aversion, as these imply that g_0 is constant or increasing with income.[66] The main case in which this assumption could fail is if relative risk aversion is increasing at large incomes. However, increasing relative risk aversion alone is not sufficient to invalidate Assumption 1: g_0 could decrease but approach some strictly positive constant at large y. Assumption 1 is unambiguously violated under constant absolute risk aversion, however: in this case e is constant over y and $g = e/y$ will become trivial as y grows large. Another case in which the assumption fails is if penalty rates and detection probabilities are sufficiently high to deter all evasion at any income level.[67] We note that our empirical results above suggest that this assumption is realistic, at least for a sizable fraction of very high-income individuals, because concealment using the types of technologies we have in mind for $a = 1$ is widespread at the top of the distribution. Assumption 1 thus allows us to construct a theoretical argument that mirrors our intuition and empirical observation.

Lemma 2. *Under Assumption 1, as y becomes arbitrarily large, $g_1(y, p_1) - g_0(y, p_1)$ converges to zero.*

[66] Assumption 1 also requires that the limit of g_0 as y tends to infinity exists, which rules out some extremely strange risk preferences and behaviors (e.g. oscillations).

[67] The rarity with which harsh penalties are actually imposed suggests that it is realistic to focus on the case where penalties do not fully deter evasion at the low detection probabilities the individual achieves by concealing evasion. See also footnote 65.

The proof of Lemma 2 and all subsequent results are in Appendix B. We solve the optimization problem (1) by first characterizing the optimal level of evasion under $a = 0$ and $a = 1$, and then comparing welfare under the two of them to determine whether the individual adopts. Lemma 2 implies that holding the probability of detection fixed at p_1, the fixed cost becomes irrelevant for behavior as y becomes large. This result helps us compare $a = 1$ and $a = 0$ for large y to determine which action the individual chooses. This comparison leads to our first main result.

Proposition 1. ***High-Income Concealment.*** *Under Assumption 1, there is a cutoff in the model $\hat{y}$ such that holding all else fixed, $y > \hat{y} \implies a = 1$ is optimal.*

Intuitively, adoption involves a trade-off between a lower probability of detection and the fixed cost. The fixed cost becomes trivial as a share of income y at large incomes. Lemma 2 states that because of this, behavior if the individual adopts the concealment technology is essentially unaffected by the fixed cost. At large incomes therefore, adoption incurs a trivial cost, but, by Assumption 1, the benefits of a lower probability of detection are non-trivial. The individual therefore adopts at sufficiently high income. Moreover, this logic applies even in the case where marginal utility u' becomes trivial for large y; covering this case makes the proof more involved than one might naively expect.

We note that Proposition 1 is not a unique cutoff rule, wherein individuals adopt *if and only if* true income exceeds some threshold. A wide parameter search of simulations of the model suggest that under constant relative risk aversion, optimal concealment does in fact follow a unique cutoff rule over y. However, we do not explicitly characterize the conditions under which such a single cutoff rule result obtains. In any case, it may be unrealistic to expect that concealment behaviors are always exclusively concentrated at the very top of the income distribution. For example, the use of cash to conceal transactions, potentially even from auditors, is generally believed to be widespread for self-employment income throughout the income distribution (see e.g. Slemrod et al., 2017). The result in Proposition 1 does however provide an explanation why in many cases, the adoption of complex and dubious sheltering strategies is concentrated at the top.

Audit Rates and Concealment. We now show that changing audit probabilities can induce the adoption of concealment. We do not condition this analysis on income for simplicity, but we discuss the results in relation to large empirical changes in audit probabilities for high-income taxpayers specifically.[68]

Proposition 2. ***Incentivizing Concealment.*** *Holding all else fixed, increasing the probability of detection under $a = 0$ will weakly increase concealment.*

Proposition 2 implies that if there is a concealment action that shields evasion against a particular type of enforcement, increasing that type of enforcement incentivizes adoption of that conceal-

[68] Note that empirical audit probabilities are based on reported income not true income; the publicly available statistics generally report audit rates by reported "total positive income" (IRS, 2020).

ment strategy. Our results suggest that broad 1040 audits like NRP random audits (and the SBSE audits with identical procedures) do not detect some types of evasion. Increasing the frequency of these types of audits could therefore incentivize adoption of sophisticated types of evasion. It also implies that increasing more sophisticated types of audits could incentivize *even more sophisticated* types of concealment, if available. Under a slightly different interpretation of the model, the proposition implies that frequent audits could incentivize the adoption of gray-area avoidance strategies that would require protracted litigation to challenge. Altogether, the fact that audits overall are relatively common at the top of the income distribution (see Figure A6a) suggests that a variety of more sophisticated concealment and dubious avoidance activities should be more prevalent at the top, all else equal.

Implications for Estimating Overall Under-Reporting We next formalize the implications of these theoretical results for the estimation of overall under-reporting and the tax gap. For simplicity, we assume that income is the only source of heterogeneity in our model; the basic point we make here does not rest on this assumption. Let $p_{D|a}$ denote the probability that a random audit detects evasion for $a \in \{0, 1\}$. We suppose that detection by random audit, like the overall probability of detection, is smaller when $a = 1$: $p_{D|a=1} < p_{D|a=0}$. If concealment is optimal at income y, so $a(y) = 1$, then evasion is detected in a random audit with probability $p_{D|a=1}$, in which case the level of evasion the individual chooses under $a = 1$ is added to the tax gap. The *estimated rate of under-reporting* conditional on true income y will be $\hat{g}(y) = a(y)p_{D|a=1} * g_1 + (1 - a(y))p_{D|a=0} * g_0$. Expressing the actual rate of under-reporting as $g = a(y)g_1 + (1 - a(y))g_0$, we have that the bias in the estimated rate of under-reporting is

$$\hat{g}(y) - g(y) = -a(y)(1 - p_{D1})g_1(y, p_1) - (1 - a(y))(1 - p_{D0})g_0(y, p_0). \tag{2}$$

We under-estimate the rate of under-reporting when random audits do not detect all evasion. More importantly, because $p_{D1} < p_{D0}$, the bias is larger when $a(y) = 1$. Our results above suggest adoption will be prevalent at high incomes, implying that this bias is particularly important at the top of the distribution. Because $g_1(y, p_1) > g_0(y, p_0)$, widespread concealment by high-income earners can even imply that the estimated rate of under-reporting decreases with income while the true rate increases with income.

7.2 Implications for Tax Administration

In Appendix C, we build a simple model to explore some implications of sophisticated evasion for optimal tax administration. The model considers the allocation of administrative resources toward audits of two types of individuals, high- and low-income. One key difference between audit allocation decisions in reality and in models of optimal tax administration (e.g. Slemrod and Yitzhaki, 2002; Keen and Slemrod, 2017) is that in reality, IRS must allocate a fixed set of resources are determined by Congress among various activities. Devoting more resources to high-income individuals requires devoting fewer resources to other activities. But in optimal tax models, no such constraint exists, and the amount of resources devoted to any particular enforcement activity is set to maximize welfare. Turning to sophistication, we argue that the adoption of sophisticated evasion strategies increases the cost of recovering revenue from high-income taxpayers. With an exogenous resource constraint, this decreases the optimal number of audits of high-income individuals, and if this effect is sufficiently strong then this can increase audits of low-income individuals. No such spillover effect occurs in optimal tax models (or equivalently when the resource constraint is set optimally), which we view as interesting given recent debates about the allocation of resources to various types of audits. Naturally, the preceding discussion holds the detection technology fixed; the presence of sophisticated evasion could increase or decrease the return to investing in specific detection technologies.

8 Conclusion

We find that substantial evasion at the top of the income distribution was undetected in individual random audits from 2006–2013. Investigating taxpayers who voluntarily declared hidden wealth or started reporting foreign bank accounts in 2009–2012, we find that audits failed to uncover offshore tax evasion. Focusing on taxpayers who earn business income through partnerships and S corporations, we find that due to the resource constraints inherent to the conduct of random audits, a large fraction of this pass-through income was not examined in the context of these audits, biasing detected evasion downward at the top. Meanwhile, random audit data on S corporations suggests that under-reporting of business income on pass-through business tax returns is substantial, five times larger than what is suggested by individual random audit data.

Theoretically, we show that modelling the choice to conceal tax evasion from auditors can explain why random audits that employ the same procedures for all taxpayers do not detect all evasion especially at the top. Empirically, we provide corrected estimates of the size and distribution of tax evasion in the United States. In our benchmark scenario, we find that under-reported income rises from about 10% of true income in the bottom 90% of the income distribution to 16% at the 99th percentile of income, where it remains constant or falls. Out of the overall top 1% under-reporting rate, at least 6 percentage points corresponds to sophisticated evasion that is seldom detected in random audits. Accounting for tax evasion increases the top 1% income share in the United States.

We stress that the limitations that we describe here pertain to NRP random audits circa 2006-2013; whether changes to audit procedures might overcome these limitations is an important question for future research. These random audit programs were not designed to estimate the tax gap for extremely high-income, high-wealth individuals. To experts who are familiar with these audits, our results may be unsurprising. However, we nevertheless view these results as important in light of an increased academic and policy interest in top income shares and tax evasion at the top. Countries around the world use random audits to estimate the tax gap (see, e.g., OECD, 2017, chapter 14). Our findings suggest that incorporating data from random audits into estimates of top-end evasion and inequality requires carefully accounting for the sophistication of top-end evasion. In this regard, statistical tools (ex. modeling the distribution of DCE-identified undetected evasion) and additional data (ex. data on offshore and pass-through level non-compliance or from more comprehensive audits of high-income taxpayers) may generate quantitatively important insights.

Our estimates are likely to be conservative with regard to the overall amount of evasion at the top. From public reporting and anecdotal evidence, it seems likely that there are other specific forms of tax evasion that have the same properties as those we examine in this paper—sophistication and concentration among high income/wealth individuals. Such forms of evasion could include, for instance, private inurement in tax-exempt organizations, and the use of offshore trusts to evade tax. Many known schemes involve entities controlled by the taxpayer in some fashion. The potential existence of more schemes than we studied above underscores the main point of our theoretical results, that we should expect sophisticated evasion to be concentrated at the top of the income and wealth distribution. More research is needed to improve estimates of noncompliance at the very top in the United States.

We identify several potentially fruitful avenues for future work. First, it would be valuable to consider the importance of sophisticated evasion and gray area avoidance strategies for optimal tax administration policies involving high-income, high-wealth taxpayers. Second, more research is needed to fully understand from an economic perspective the gray area between avoidance and evasion, a line which can be blurry at the top of the income distribution and for large corporations. Third, future research could consider *strategic* interactions between the tax authority and high-income individuals involving sophisticated evasion. Finally, future work could consider the implications of our work for white-collar, financial crime more broadly.

TABLE 1: SHARES OF UNREPORTED INCOME, 2008-2013, IN % OF TOTAL UNREPORTED INCOME

	Exam-corrected No DCE No sophisticated	Exam-corrected With micro DCE (2019) No sophisticated	Exam-corrected With MA-DCE No sophisticated	Exam-corrected No DCE With sophisticated	Exam-corrected With micro DCE (2019) With sophisticated	Exam-corrected With MA-DCE With sophisticated
P0-10	-1.3	0.9	-0.2	-0.8	0.7	-0.1
P10-20	1.8	1.0	1.6	1.2	0.9	1.4
P20-30	3.3	2.1	2.9	2.2	1.9	2.5
P30-40	4.9	3.3	3.8	3.3	2.9	3.3
P40-50	6.0	5.2	4.6	4.0	4.6	4.0
P50-60	6.9	6.4	5.3	4.7	5.6	4.7
P60-70	9.8	8.4	7.4	6.6	7.4	6.5
P70-80	11.8	11.0	9.3	8.1	9.6	8.2
P80-90	15.4	14.9	12.7	11.0	13.1	11.3
P90-95	12.9	12.1	10.8	9.7	10.7	9.8
P95-99	15.5	19.3	16.8	15.0	18.1	16.3
P99-99.5	4.7	5.1	5.6	6.1	5.6	6.0
P99.5-99.9	5.0	6.2	7.8	10.8	8.3	9.6
P99.9-P99.95	0.9	1.0	2.2	3.5	2.1	3.0
P99.95-P99.99	1.5	2.0	3.7	6.2	4.0	5.3
P99.99-100	0.9	1.0	5.7	8.3	4.4	8.1
Top 1%	13.0	15.3	24.9	34.9	24.4	31.9

Note: This table reports the distribution of unreported income across income groups, for different measures of unreported income. Throughout, tax units are ranked by their estimated true income (equal to reported income plus estimated unreported income). The first column shows the distribution of unreported income detected in NRP exams with no adjustment for undetected evasion (in particular, without including evasion identified by DCE). The second and third column shows the distribution of unreported income including DCE-identified evasion using either the DCE2019 micro or baseline MA-DCE specifications. The fourth column shows the distribution of unreported income detected in the NRP without DCE-identified evasion, but adds our benchmark estimate of sophisticated evasion. The final two columns shows the distribution of unreported income incorporating both sophisticated evasion and DCE-identified evasion.

TABLE 2: SHARES OF TRUE INCOME, 2008-2013, IN % OF TOTAL INCOME

	Reported	Exam-corrected No DCE No sophisticated	Exam-corrected With micro DCE (2019) No sophisticated	Exam-corrected With MA-DCE No sophisticated	Exam-corrected No DCE With sophisticated	Exam-corrected With micro DCE (2019) With sophisticated	Exam-corrected With MA-DCE With sophisticated
P0-10	-1.3	-1.0	-0.6	-0.9	-0.9	-0.6	-0.8
P10-20	1.4	1.5	1.5	1.5	1.4	1.5	1.5
P20-30	2.4	2.5	2.6	2.5	2.5	2.6	2.5
P30-40	3.5	3.6	3.7	3.6	3.5	3.7	3.5
P40-50	4.8	4.9	5.0	4.8	4.8	4.9	4.8
P50-60	6.4	6.5	6.5	6.3	6.3	6.4	6.2
P60-70	8.5	8.5	8.6	8.4	8.4	8.4	8.3
P70-80	11.5	11.5	11.4	11.2	11.2	11.2	11.1
P80-90	16.3	16.2	16.1	15.9	15.9	15.8	15.6
P90-95	11.8	11.7	11.6	11.5	11.5	11.4	11.4
P95-99	15.6	15.5	15.5	15.6	15.5	15.4	15.6
P99-99.5	4.0	4.0	3.9	4.1	4.1	4.0	4.2
P99.5-99.9	6.3	6.1	6.0	6.4	6.5	6.2	6.6
P99.9-P99.95	1.8	1.8	1.7	1.8	1.9	1.8	1.9
P99.95-P99.99	2.9	2.8	2.7	3.0	3.1	2.9	3.2
P99.99-100	4.1	4.0	3.8	4.3	4.4	4.1	4.6
Top 1%	19.1	18.7	18.1	19.5	19.9	19.2	20.4%

Notes: This table reports estimates of the distribution of true market income (defined as total income reported on form 1040 minus Social Security benefits, unemployment insurance benefits, alimony, and state refunds) for different measures of unreported income. The first column shows the distribution of reported income (i.e., with no adjustment for evasion). In the second column, under-reported income only includes what is detected in NRP random audit data with no adjustment of any kind (in particular without DCE-identified under-reported income). In the third and fourth column, under-reported income is equal to DCE-adjusted NRP unreported income, using either the DCE2019 micro specification or the baseline MA-DCE specification. In the fifth column, under-reported income is equal to exam-detected unreported income (without DCE-identified unreported income) plus sophisticated evasion. The last two columns estimate the distribution of true income including both DCE-identified unreported income and sophisticated evasion.

References

Allingham, M. G. and Sandmo, A. (1972). Income tax evasion: A theoretical analysis. *Journal of Public Economics*, 1(3-4):323–338.

Alstadaeter, A., Johannesen, N., and Zucman, G. (2018). Who owns the wealth in tax havens? Macro evidence and implications for global inequality. *Journal of Public Economics*, 162:89 – 100.

Alstadsaeter, A., Johannesen, N., and Zucman, G. (2019). Tax Evasion and Inequality. *American Economic Review*, 109(6):2073–2103.

Andreoni, J., Erard, B., and Feinstein, J. (1998). Tax Compliance. *Journal of Economic Literature*, 36(2):818–860.

Auten, G. and Splinter, D. (2019). Income inequality in the united states: Using tax data to measure long-term trends. Working paper.

Blumenthal, M., Christian, C., and Slemrod, J. (1998). The determinants of income tax compliance: Evidence from a controlled experiment in minnesota. NBER working paper no. 6575.

Brown, R. E. and Mazur, M. J. (2003). IRS's comprehensive approach to compliance measurement. *National Tax Journal*, 56(3).

Carrillo, P., Pomeranz, D., and Singhal, M. (2017). Dodging the taxman: Firm misreporting and limits to tax enforcement. *American Economic Journal: Applied Economics*, 9(2):144–64.

Casi, E., Spengel, C., and Stage, B. M. (2020). Cross-border tax evasion after the common reporting standard: Game over? *Journal of Public Economics*, 190:104240.

Cooper, M., McClelland, J., Pearce, J., Prisinzano, R., Sullivan, J., Yagan, D., Zidar, O., and Zwick, E. (2016). Business in the United States: Who Owns it, and How Much Tax Do They Pay? *Tax Policy and the Economy*, 30(1):91–128.

De Simone, L., Lester, R., and Markle, K. (2020). Transparency and tax evasion: Evidence from the foreign account tax compliance act (fatca). *Journal of Accounting Research*, 58(1):105–153.

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

DOJ (2021). Offshore compliance initiative news indictments, pleas, sentencings, and other developments. Archive, accessed june 2021, available online at https://www.justice.gov/tax/offshore-compliance-initiative.

Erard, B. and Feinstein, J. (2011). The individual income reporting gap: What we see and what we don't. 2011 irs-tpc research conference, publication 1500 (rev 4-2012), Washington, DC.

Feinstein, J. S. (1991). An Econometric Analysis of Income Tax Evasion and Its Detection. *The RAND Journal of Economics*, 22(1):14–35.

GAO (1995). Tax administration: IRS' partnership compliance activities could be improved. Letter report, 06/16/95, gao/ggd-95-151.

Guttentag, J. and Avi-Yonah, R. S. (2005). Closing the international tax gap.

Guyton, J., Langetieg, P., Reck, D., Risch, M., and Zucman, G. (2023). The distribution of undetected under-reported income identified by detection-controlled estimation methods. Unpublished working paper, included in the online appendix to this paper.

IRS (2007). Tax year 2001 federal tax gap overview. https://www.irs.gov/pub/irs-soi/01rastg07map.pdf.

IRS (2016a). Federal Tax Compliance Research: Tax Gap Estimates for Tax Years 2008-2010. Publication 1415 (rev. 5-2016), Washington, DC.

IRS (2016b). Transaction of interest – section 831(b) micro-captive transactions. *Internal Revenue Bulletin 2016-46*, notice 2016-66.

IRS (2019). Federal Tax Compliance Research: Tax Gap Estimates for Tax Years 2011-2013. Publication 1415 (rev. 9-2019), Washington, DC.

IRS (2020). Internal revenue service data book 2019. Publication 55-b (rev. 06-2020), Washington, DC.

IRS (2022). Internal revenue service data book 2021. Publication 55-b (rev. 2021), Washington, DC.

Johannesen, N., Langetieg, P., Reck, D., Risch, M., and Slemrod, J. (2020). Taxing hidden wealth: The consequences of us enforcement initiatives on evasive foreign accounts. *American Economic Journal: Economic Policy*, 12(3):312–46.

Johns, A. and Slemrod, J. (2010). The Distribution of Income Tax Noncompliance. *National Tax Journal*, 63(3):397–418.

Johns, D. (2009). Preliminary results of the 2003/2004 national research program s corporation underreporting study. presentation, 2009 IRS research conference. Technical report.

Joulfaian, D. (2000). Corporate income tax evasion and managerial preferences. *Review of Economics and Statistics*, 82(4):698–701.

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

Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S., and Saez, E. (2011). Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark. *Econometrica*, 79(3):651–692.

Kleven, H. J., Kreiner, C. T., and Saez, E. (2016). Why can modern governments tax so much? an agency model of firms as fiscal intermediaries. *Economica*, 83(330):219–246.

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

Luttmer, E. F. and Singhal, M. (2014). Tax morale. *Journal of economic perspectives*, 28(4):149–68.

OECD (2000). Towards global tax co-operation. Report to the 2000 ministerial council meeting and recommendations by the committee on fiscal affairs.

OECD (2017). *Tax Administration 2017*.

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

Piketty, T., Saez, E., and Zucman, G. (2018). Distributional national accounts: methods and estimates for the united states. *Quarterly Journal of Economics*, 133(2):553–609.

Reck, D., Risch, M., and Zucman, G. (2021). Response to a comment by Auten and Splinter on "tax evasion at the top of the income distribution: Theory and evidence". Mimeo, available at https://www.danreck.com/s/asresponse_full.pdf.

Saez, E. and Zucman, G. (2016). Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data. *The Quarterly Journal of Economics*, 131(2):519–578.

Saez, E. and Zucman, G. (2020). Trends in US Income and Wealth Inequality: Revising after the Revisionists. NBER working paper no. 27921.

Slemrod, J., Collins, B., Hoopes, J. L., Reck, D., and Sebastiani, M. (2017). Does credit-card information reporting improve small-business tax compliance? *Journal of Public Economics*, 149:1–19.

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

Smith, M., Yagan, D., Zidar, O., and Zwick, E. (2019a). Capitalists in the twenty-first century. *Quarterly Journal of Economics*, 134(4):1675–1745.

Smith, M., Zidar, O., and Zwick, E. (2019b). Top wealth in the united states: New estimates and implications for taxing the rich. *Unpublished*.

Swiss National Bank (2007). Banks in switzerland, 2006 edition. annual report of the swiss central bank, Zurich.

TIGTA (2020). High-income nonfilers owing billions of dollars are not being worked by the internal revenue service. Treasury inspector general for tax administration publication 2020-30-015.

United States Senate (2008). Tax Haven Banks and U.S. Tax compliance. Staff report of the permanent subcommittee on investigations, Washington, DC.

United States Senate (2014). Offshore Tax Evasion: The Effort to Collect Unpaid Taxes on Billions in Hidden Offshore Accounts. Technical report, Washington, DC.

United States Senate (2020). Syndicated conservation-easement transactions. Investigative report of the senate finance committee, Washington, DC.

Yitzhaki, S. (1974). A note on 'income tax evasion: A theoretical analysis'. *Journal of Public Economics*, 3(2):201–202.

Zucman, G. (2013). The Missing Wealth of Nations: Are Europe and the U.S. net Debtors or net Creditors? *The Quarterly Journal of Economics*, 128(3):1321–1364.

Zucman, G. (2014). Taxing Across Borders: Tracking Personal Wealth and Corporate Profits. *Journal of Economic Perspectives*, 28(4):121–148.

"Tax Evasion at the Top of the Income Distribution: Theory and Evidence" Appendix

John Guyton, Patrick Langetieg, Daniel Reck,
Max Risch, and Gabriel Zucman

This appendix contains the following supplementary material:

- Appendix A contains additional empirical results referenced in the main text.
- Appendix B proves the results in Section 7 of the main text.
- Appendix C lays out a model in which we explore the implications of some of our results for tax administration.

A Additional Results

TABLE A1: INCOME UNDER-REPORTING DETECTED IN NRP RANDOM AUDITS: DECOMPOSITION BY INCOME TYPE, TAX YEARS 2006–2013

	Full Population				Top 1%			
	Total income of this type/ Total income (%)	Total under-reported income of this type/ Total under-reported income (%)	Total under-reported income of this type/ Total income (%)	Total under-reported income of this type/ Total income of this type (%)	Total income of this type/ Total income (%)	Total under-reported income of this type/ Total under-reported income (%)	Total under-reported income of this type/ Total income (%)	Total under-reported income of this type/ Total income of this type (%)
Capital Gains	5.9	7.1	0.28	4.8	21.4	18.8	0.43	2.0
Dividends	3.9	2.8	0.11	2.8	8.6	3.3	0.09	0.9
Interest	1.9	0.7	0.03	1.5	3.0	2.0	0.05	1.6
Line 21 Other Income	0.2	11.9	0.47	242.3	2.6	8.5	0.19	7.5
Partnerships and S Corp	5.6	6.5	0.26	4.6	21.7	18.8	0.43	2.0
Rental	0.7	8.9	0.35	48.1	1.6	5.4	0.12	7.9
Schedule C	5.3	49.4	1.95	36.7	4.2	34.9	0.79	18.7
Wages	72.3	3.5	0.14	0.2	38.1	2.9	0.07	0.2
Other	4.1	9.2	0.37	0.1	-1.1	5.3	0.10	-0.1
Total	100.0	100.0	3.95		100.0	100.0	2.27	

Note: This table describes the composition of detected under-reported income in the 2006–2013 NRP data, before any correction for undetected noncompliance (in particular before any DCE correction). The NRP shares of each type of income in total income are similar to income shares we observe in Statistics of Income (SOI) data, but the NRP shares are built using corrected income here. Consequently the largest differences with SOI income shares are observed for types of income with significant detected evasion. Note that "Form 1040 Other Income" in Figure 1a is referred to as "Line 21 Other Income" here, as this item appears on Line 21 of the Form 1040, while the residual "Other" category in the penultimate row refers to all other components of income. We note that the estimated rates of under-reporting by type of income in the fourth column are well in excess of 100% for Line 21 income. This can occur because Line 21 income can be negative; large negative values are common at the bottom of the income distribution because of net operating loss carryforwards or carrybacks from pass-through businesses. Large corrections to line 21 are typically disallowed loss carryforwards or carrybacks.

Table A2: Reranking: Detected Under-Reporting by Reported and Corrected Market Income, as a % of Total Detected Under-Reporting, TY2006–2013

	P0-10	P10-20	P20-30	P30-40	P40-50	P50-60	P60-70	P70-80	P80-90	P90-95	P95-99	P99-100	Row Total
P0-10	4.9	1.4	1.3	1.4	1.2	1.3	1.4	1.8	2.0	1.4	1.5	1.2	21.0
P10-20	-0.7	0.6	1.1	1.1	0.8	0.7	0.7	0.5	0.4	**	**	**	5.4
P20-30	-1.0	-0.4	0.8	1.7	1.7	1.0	1.1	0.7	0.8	0.4	**	**	7.3
P30-40	-0.7	-0.4	-0.2	1.0	1.9	1.4	1.1	1.1	0.8	0.7	**	**	7.7
P40-50	-0.2	**	-0.1	-0.2	0.9	2.1	1.5	1.3	1.1	0.7	0.3	**	7.4
P50-60	**	**	**	-0.1	-0.1	1.3	2.5	1.5	1.4	0.9	0.8	**	8.6
P60-70	**	**	**	**	0.0	-0.1	1.8	3.0	2.1	1.0	0.9	**	9.2
P70-80	**	**	**	**	**	-0.1	-0.2	2.4	3.5	1.6	1.4	**	8.5
P80-90	**	**	**	**	**	**	-0.1	-0.3	3.6	3.1	2.0	**	8.8
P90-95	**	**	**	**	**	**	**	**	-0.2	2.4	2.8	0.5	5.4
P95-99	**	**	**	**	**	**	**	**	**	-0.1	5.0	2.3	6.8
P99-100	**	**	**	**	**	**	**	**	**	**	-0.2	4.5	4.1
Col. Total	1.4	0.7	2.8	4.9	6.2	7.6	9.7	11.8	15.6	12.2	15.5	11.5	100.0

Note: This table reports the share of under-reporting in cells of both reported and corrected income using the NRP data. Each entry is total under- or over-reporting in that cell, scaled by total net under-reporting in the full population. The final row presents column totals and the final column presents row totals. Output is suppressed for cells containing fewer than 10 observations; the row and column totals include the suppressed amounts. We note that rounding issues cause tiny discrepancies between the column totals and the same information reported in Table A7. The main lesson we take away from this table comes from the P0-P10 part of the distribution: from the totals, we find that 21% of total under-reporting locates in P0-P10 by reported income, but just 1.4% of total (net) under-reporting locates in the bottom bin by corrected market income. If we condition on having reported income in P0-P10, we find that about 23% of under-reporting in this bin (=4.9/21.0) remains in the bottom bin after re-ranking by corrected income, while the remaining 76% moves upward in the distribution due to re-ranking effects. After re-ranking, the 4.9% of under-reporting that remains in P0-P10 is joined by *over-reporting* from individuals who were previously were in a higher part of the income distribution, bringing total net under-reporting in P0-P10 by corrected income down to 1.4% of all under-reporting.

Table A3: Offshore evasion scenarios

Parameter	Lower-bound scenario	Preferred scenario	Upper-bound scenario
Amount of U.S. offshore wealth (in billion $)	750	1,058	1,500
Fraction of offshore wealth concealed	85%	95%	100%
Rate of return on offshore wealth	4.65 %	6%	11%
Distribution of offshore wealth	FBAR	Average of FBAR and Nordic	Nordic
Average Marginal Tax Rate	20%	25%	30%

Note: This table summarizes the five sets of assumptions about the amount and distribution of offshore income made in our three different scenarios discussed in Section 3.3.

Table A4: Entity-Level Income Under-Reporting Detected in Random Audits of S Corporations: Decomposition by Line Item, TY2003–2004

	Total under-reporting attributable to this line / total under-reporting of net S corp income	Total under-reporting attributable to this line / total corrected net S corp income	Total under-reporting attributable to this line / total true amount of this line
Gross receipts or sales	26.27	5.11	0.32
Returns and allowances	-0.06	-0.01	-0.12
Cost of goods sold (-)	13.84	2.69	0.26
Gross profit	40.26	7.83	1.35
Net gain	0.62	0.12	4.41
Other income	4.60	0.89	2.00
Total income	45.50	8.85	1.41
Compensation of officers (-)	-24.96	-4.86	-7.15
Salaries and wages (-)	1.42	0.28	0.18
Repairs and maintenance (-)	2.82	0.55	6.33
Bad debts (-)	0.90	0.17	6.98
Rents (-)	3.07	0.60	1.88
Taxes and licenses (-)	1.28	0.25	0.92
Interest (-)	1.92	0.37	2.86
Depreciation not claimed (-)	5.33	1.04	3.64
Depletion (-)	0.14	0.03	20.30
Advertising (-)	1.15	0.22	1.47
Pension, profit-sharing, etc., plans (-)	0.64	0.12	2.04
Employee benefit programs (-)	1.10	0.21	2.43
Other deductions (-)	34.75	6.76	3.95
Total deductions (-)	29.51	5.74	1.09
Ordinary business income or loss	75.01	14.59	14.59
Ordinary business income excl. officer compensation	100.0	19.45	19.45

Note: This table presents a decomposition of entity-level under-reporting on the tax returns of S corporations. Each line of the table corresponds to a line on the Form 1120-S. Line items in bold in the first column are aggregations of other lines. The first column reports each line item's share of overall mis-reporting in S corporations according to our preferred definition (exlcuding officer compensation, see the main text for details). The second column reports each line's additive contribution to the 19.45% under-reporting rate for overall mis-reporting rate. The third column reports the mis-reporting rate for each line, defined as the amount under-reporting due to that line item as a share of the line item total, i.e. a rate of mis-reporting for each line. Deductions are marked with a (-) in the first column. The entries in the rows corresponding to deductions are positive when on net, mis-reporting on that line *increases* under-reporting of net business income. For example, officer compensation tends to be under-reported so before we exclude it in our preferred measure of overall under-reporting, it has a negative contribution to under-reporting of net business income.

TABLE A5: PASS-THROUGH EVASION SCENARIOS

Parameter	Lower-bound scenario	Preferred scenario	Upper-bound scenario
Under-reporting of net business income (%)	S corp NRP	20%	28%
Under-reporting of net business income (%)	0	5%	10%
Under-reporting of net business income (%)	0	3%	6%
Distribution of unreported income	S corp NRP	Reported passthrough income	Reported passthrough income

Note: This table summarizes the four sets of assumptions about the amount and distribution of passthrought income made in our three different scenarios discussed in Section 4.

TABLE A6: INCOME UNDER-REPORTING BASED ON RANDOM AUDIT DATA INCLUDING DCE-IDENTIFIED EVASION: DECOMPOSITION BY INCOME TYPE, TAX YEARS 2006–2012

	Total income of this type/ Total income (%)	Total under-reported income of this type/ Total under-reported income (%)	Total under-reported income of this type/ Total income (%)	Total under-reported income of this type/ Total income of this type (%)
Capital Gains	5.4	11.9	1.27	23.4
Dividends	2.1	1.0	0.11	5.2
Interest	1.5	0.3	0.03	2.1
Line 21 Other Income	0.9	10.2	1.09	126.1
Partnerships and S Corp	5.6	8.0	0.85	15.3
Rental	1.5	10.6	1.13	75.2
Schedule C	7.6	42.3	4.52	59.8
Wages	66.3	5.4	0.58	0.9
Other	9.3	10.2	1.09	11.8
Total	100.0	100.0	10.7	

Note: This table describes the composition of detected under-reported income in the 2006–2013 NRP data, including DCE-identified undetected under-reporting. The NRP shares of each type of income in total income are similar to income shares we observe in Statistics of Income (SOI) data, but the NRP shares are built using DCE-adjusted income here. Consequently the largest differences with SOI income shares are observed for types of income with significant detected evasion. Note that "Form 1040 Other Income" in Figure 1a is referred to as "Line 21 Other Income" here, as this item appears on Line 21 of the Form 1040, while the residual "Other" category in the penultimate row refers to all other components of income. We note that the estimated rates of under-reporting by type of income in the fourth column exceeds 100% for Line 21 income. This can occur because Line 21 income can be negative; large negative values are common at the bottom of the income distribution because of net operating loss carryforwards or carrybacks from pass-through businesses. Large corrections to line 21 are typically disallowed loss carryforwards or carrybacks.

TABLE A7: SHARES OF UNREPORTED INCOME, 2006-2013, IN % OF TOTAL UNREPORTED INCOME

	Exam-corrected No DCE No sophisticated	Exam-corrected With MA-DCE No sophisticated	Exam-corrected No DCE With sophisticated	Exam-corrected With MA-DCE With sophisticated
P0-10	1.4	0.7	1.0	0.8
P10-20	0.7	0.8	0.5	0.7
P20-30	2.8	2.4	1.9	2.1
P30-40	4.9	3.7	3.3	3.2
P40-50	6.2	4.5	4.1	3.9
P50-60	7.6	5.6	5.1	4.9
P60-70	9.7	7.5	6.6	6.6
P70-80	11.9	9.7	8.2	8.6
P80-90	15.6	13.1	11.1	11.7
P90-95	12.2	10.8	9.2	9.8
P95-99	15.7	17.3	15.2	16.8
P99-99.5	3.9	5.5	5.6	5.9
P99.5-99.9	5.1	8.0	10.9	9.7
P99.9-P99.95	1.1	2.2	3.6	3.0
P99.95-P99.99	0.7	3.2	5.7	4.8
P99.99-100	0.6	5.0	8.1	7.5
Top 1%	11.4	23.8	33.9	30.9

Note: This table reports the distribution of unreported income across income groups, for different measures of unreported income, for tax years 2006–2013. Tax units are ranked by their estimated true income (equal to reported income plus estimated unreported income). The first column shows the distribution of unreported income detected in NRP exams with no adjustment for undetected evasion. The second column shows the distribution of unreported income including under-reporting identified by DCE methods using our baseline MA-DCE specification. We are unable to implement the micro DCE2019 methodology in this range of tax years. The third column shows the distribution of unreported income detected in the NRP without any DCE adjustments, but adds our benchmark estimate of sophisticated evasion. The last column shows the distribution of unreported income including both sophisticated evasion and the MA-DCE adjustment. Table 1 reports analogous results for tax years 2008–2013, including results employing micro DCE2019.

TABLE A8: SHARES OF TRUE INCOME, 2006-2013, IN % OF TOTAL INCOME

	NRP Before exam	NRP After exam No DCE No sophisticated	NRP After exam With MA-DCE No sophisticated	NRP After exam No DCE Add sophisticated	Our benchmark
P0-10	-2.6	-2.1	-1.9	-2.0	-1.9
P10-20	1.0	1.0	1.0	1.0	1.0
P20-30	2.1	2.1	2.1	2.1	2.1
P30-40	3.2	3.4	3.3	3.3	3.3
P40-50	4.7	4.8	4.7	4.7	4.6
P50-60	6.4	6.5	6.4	6.4	6.3
P60-70	8.6	8.7	8.5	8.5	8.4
P70-80	11.7	11.6	11.4	11.4	11.3
P80-90	16.6	16.4	16.1	16.2	15.9
P90-95	12.0	11.8	11.7	11.6	11.6
P95-99	16.1	16.0	16.1	15.9	16.1
P99-99.5	4.3	4.2	4.4	4.3	4.4
P99.5-99.9	6.7	6.5	6.7	6.9	7.0
P99.9-P99.95	2.0	1.9	2.0	2.0	2.1
P99.95-P99.99	3.2	3.0	3.1	3.3	3.3
P99.99-100	4.2	4.1	4.3	4.5	4.6
Top 1%	20.3	19.8	20.5	21.0	21.4

Note: This table reports the distribution of true market income (defined as total income reported on form 1040 minus Social Security benefits, unemployment insurance benefits, alimony, and state refunds) for different measures of unreported income. The first column shows the distribution of reported income (i.e., with no adjustment for evasion). In the second column, under-reported income only includes what is detected in raw random audits with no adjustment of any kind (in particular no DCE adjustment). In the third column, under-reported income is equal to DCE-adjusted NRP unreported income. In the fourth column, under-reported income is equal to exam-detected under-reporting plus sophisticated evasion, without DCE-identified evasion. The last column shows the distribution of true income in our benchmark scenario combining the benchmark MA-DCE adjustment and sophisticated evasion.

TABLE A9: INCOME SHARES IN TY2008-2013 NRP DATA AND IN TY2001 NRP DATA

	2008–2013 Before exam	2008–2013 After exam No DCE	2008–2013 After exam MA-DCE	2008–2013 With sophisticated MA-DCE	2008–2013 After exam Micro DCE (2019)	2008–2013 With sophisticated Micro DCE (2019)	2001 Before exam	2001 After exam Micro DCE (2001)
P0-10	-1.3	-1.0	-0.9	-0.8	-0.6	-0.6	0.1	0.3
P10-20	1.4	1.5	1.5	1.5	1.5	1.5	1.6	1.6
P20-30	2.4	2.5	2.5	2.5	2.6	2.6	2.7	2.7
P30-40	3.5	3.6	3.6	3.5	3.7	3.7	3.9	3.9
P40-50	4.8	4.9	4.8	4.8	5.0	4.9	5.2	5.2
P50-60	6.4	6.5	6.3	6.2	6.5	6.4	6.8	6.7
P60-70	8.5	8.5	8.4	8.3	8.6	8.4	8.9	8.8
P70-80	11.5	11.5	11.2	11.1	11.4	11.2	11.7	11.5
P80-90	16.3	16.2	15.9	15.6	16.1	15.8	16	15.6
P90-95	11.8	11.7	11.5	11.4	11.6	11.4	11	10.9
P95-99	15.6	15.5	15.6	15.6	15.5	15.4	14.4	14.9
P99-99.5	4.0	4.0	4.1	4.2	3.9	4.0	3.7	3.8
Top 0.5%	15.1	14.7	15.4	16.3	14.2	15.1	14.1	14

Note: This table reports the distribution of income in the 2008–2013 NRP data studied in this paper and in the 2001 NRP data as reported in Johns and Slemrod (2010, Table 5). Tax units are ranked by their estimated true income (equal to reported income plus estimated under-reported income). Income is Adjusted Gross Income (AGI) in Johns and Slemrod (2010) and market income in our series (defined as total income reported on form 1040 minus Social Security benefits, unemployment insurance benefits, alimony, and state refunds).

TABLE A10: INCOME SHARES IN TY2006-2013 NRP DATA AND IN TY2001 NRP DATA

	2006–2013 Before exam	2006–2013 After exam No DCE	2006–2013 After exam MA-DCE	2006–2013 with sophisticated MA-DCE	2001 Before exam	2001 After exam Micro DCE (2001)
P0-10	-2.6	-2.1	-1.9	-1.9	0.1	0.3
P10-20	1.0	1.0	1.0	1.0	1.6	1.6
P20-30	2.1	2.1	2.1	2.1	2.7	2.7
P30-40	3.2	3.4	3.3	3.3	3.9	3.9
P40-50	4.7	4.8	4.7	4.6	5.2	5.2
P50-60	6.4	6.5	6.4	6.3	6.8	6.7
P60-70	8.6	8.7	8.5	8.4	8.9	8.8
P70-80	11.7	11.6	11.4	11.3	11.7	11.5
P80-90	16.6	16.4	16.1	15.9	16	15.6
P90-95	12.0	11.8	11.7	11.6	11	10.9
P95-99	16.1	16.0	16.1	16.1	14.4	14.9
P99-99.5	4.3	4.2	4.4	4.4	3.7	3.8
Top 0.5%	16.0	15.6	16.2	17.0	14.1	14

Note: This table reports the distribution of income in the 2006–2013 NRP data studied in this paper and in the 2001 NRP data as reported in Johns and Slemrod (2010, Table 5). Tax units are ranked by their estimated true income (equal to reported income plus estimated under-reported income). Income is Adjusted Gross Income (AGI) in Johns and Slemrod (2010) and market income in our series (defined as total income reported on form 1040 minus Social Security benefits, unemployment insurance benefits, alimony, and state refunds). Series in columns 3 and 4 use our baseline MA-DCE; column 6 uses micro DCE2001 on 2001 NRP data. We are unable to implement micro DCE2019 for the full 2006–2013 period; see also Table A9.

TABLE A11: COMPARING DCE SPECIFICATIONS: UNDER-REPORTED INCOME AS A % OF TOTAL UNDER-REPORTED INCOME - TY2008-2013

	Exam-corrected income, no sophisticated	Exam-corrected income, add sophisticated	Baseline MA-DCE (Exam-corrected income share)	Micro DCE2019	MA-DCE Exam-detected under-rep share	MA-DCE Modified reported income share	MA-DCE Reported income share
P0-10	-1.3	-0.8	-0.1	0.7	0.5	0.0	0.0
P10-20	1.8	1.2	1.4	0.9	1.5	1.1	1.6
P20-30	3.3	2.2	2.5	1.9	2.6	2.0	2.8
P30-40	4.9	3.3	3.3	2.9	4.0	2.5	3.1
P40-50	6.0	4.0	4.0	4.6	4.7	3.0	3.5
P50-60	6.9	4.7	4.7	5.6	5.4	3.7	4.3
P60-70	9.8	6.6	6.5	7.4	8.0	5.2	5.9
P70-80	11.8	8.1	8.2	9.6	9.7	6.8	7.5
P80-90	15.4	11.0	11.3	13.1	13.0	10.1	11.4
P90-95	12.9	9.7	9.8	10.7	11.4	8.5	9.4
P95-99	15.5	15.0	16.3	18.1	15.1	16.6	17.7
P99-99.5	4.7	6.1	6.0	5.6	5.4	6.9	6.3
P99.5-99.9	5.0	10.8	9.6	8.3	8.1	12.8	10.2
P99.9-P99.95	0.9	3.5	3.0	2.1	2.2	4.2	3.2
P99.95-P99.99	1.5	6.2	5.3	4.0	3.7	7.0	5.3
P99.99-100	0.9	8.3	8.1	4.4	4.7	9.6	7.9
Top 1%	13.0	34.9	31.9	24.4	24.0	40.5	32.9

Note: This table illustrates the sensitivity of the results in Table 1 to assumptions about the location in the income distribution of evasion identified by DCE methods. The first three columns appear in Table 1 and are shown here for comparison purposes. The fourth column implements the micro DCE2001 method used on 2001 NRP data by Johns and Slemrod (2010). We then report estimates using alternative macro allocation methods considered in Figure A14 in the final three columns. See Guyton et al. (2023) for additional details on DCE specifications.

Table A12: Comparing DCE Specifications: True Income as a % of Total True Income - TY2008-2013

	Reported income (no sophisticated)	Exam-corrected income, no sophisticated	Exam-corrected income, add sophisticated	Baseline MA-DCE (Exam-corrected income share)	Micro DCE2019	MA-DCE Exam-detected under-rep share	MA-DCE Modified reported income share	MA-DCE Reported income share
P0-10	-1.3	-1.0	-0.9	-0.8	-0.6	-0.8	-0.8	-0.8
P10-20	1.4	1.5	1.4	1.5	1.5	1.5	1.4	1.5
P20-30	2.4	2.5	2.5	2.5	2.6	2.5	2.4	2.5
P30-40	3.5	3.6	3.5	3.5	3.7	3.6	3.4	3.5
P40-50	4.8	4.9	4.8	4.8	4.9	4.8	4.6	4.7
P50-60	6.4	6.5	6.3	6.2	6.4	6.3	6.1	6.2
P60-70	8.5	8.5	8.4	8.3	8.4	8.4	8.1	8.2
P70-80	11.5	11.5	11.2	11.1	11.2	11.3	10.9	11.0
P80-90	16.3	16.2	15.9	15.6	15.8	15.8	15.5	15.6
P90-95	11.8	11.7	11.5	11.4	11.4	11.6	11.2	11.3
P95-99	15.6	15.5	15.5	15.6	15.4	15.4	15.6	15.7
P99-99.5	4.0	4.0	4.1	4.2	4.0	4.1	4.3	4.2
P99.5-99.9	6.3	6.1	6.5	6.6	6.2	6.4	7.0	6.7
P99.9-P99.95	1.8	1.8	1.9	1.9	1.8	1.8	2.1	2.0
P99.95-P99.99	2.9	2.8	3.1	3.2	2.9	3.0	3.4	3.2
P99.99-100	4.1	4.0	4.4	4.6	4.1	4.2	4.8	4.6
Top 1%	19.1	18.7	19.9	20.4	19.2	19.5	21.5	20.6%

Note: This table illustrates the sensitivity of the results in Table 2 to assumptions about the location in the income distribution of evasion identified by DCE methods. The first four columns appear in Table 2 and are shown here for comparison purposes. The fifth column implements the micro DCE2001 method used on 2001 NRP data by Johns and Slemrod (2010). We then report estimates using alternative macro allocation methods considered in Figure A14 in the final three columns. See Guyton et al. (2023) for additional details on DCE specifications. Naturally, methods allocating less undetected evasion from DCE at the top imply smaller estimates of the top 1% income share. However, accounting for sophisticated evasion has a substantial impact regardless of the DCE allocation, and we observe that accounting for evasion, including sophisticated evasion, increases the top 1 share relative to reported incomes for any of the allocation methods we consider.

TABLE A13: COMPARING DCE SPECIFICATIONS: UNDER-REPORTED INCOME AS A % OF TOTAL UNDER-REPORTED INCOME - TY2006-2013

Bin	Exam-corrected income, no sophisticated	Exam-corrected income, add sophisticated	Baseline MA-DCE (Exam-corrected income share)	Micro DCE2001	MA-DCE Exam-detected under-rep share	MA-DCE Modified reported income share	MA-DCE Reported income share
P0-10	1.4	1.0	0.8	-0.3	0.8	2.8	0.8
P10-20	0.7	0.5	0.7	0.2	1.2	0.6	0.8
P20-30	2.8	1.9	2.1	0.9	2.5	2.3	1.8
P30-40	4.9	3.3	3.2	1.4	3.0	3.9	2.5
P40-50	6.2	4.1	3.9	2.0	3.5	5.0	3.1
P50-60	7.6	5.1	4.9	3.2	4.4	5.9	3.9
P60-70	9.7	6.6	6.6	4.6	5.7	7.7	5.2
P70-80	11.9	8.2	8.6	7.2	7.4	9.7	6.8
P80-90	15.6	11.1	11.7	11.6	11.2	13.2	10.1
P90-95	12.2	9.2	9.8	11.4	9.0	11.0	8.3
P95-99	15.7	15.2	16.8	24.8	17.7	15.5	16.7
P99-99.5	3.9	5.6	5.9	7.9	6.2	4.6	6.6
P99.5-99.9	5.1	10.9	9.7	12.3	10.7	8.4	12.8
P99.9-P99.95	1.1	3.6	3.0	3.3	3.4	2.3	4.3
P99.95-P99.99	0.7	5.7	4.8	4.4	5.3	3.0	6.7
P99.99-100	0.6	8.1	7.5	5.3	8.0	4.2	9.5
Top 1%	11.4	33.9	30.9	33.1	33.6	22.4	40.0

Note: This table illustrates the sensitivity of the results in Table A7 to assumptions about the location in the income distribution of evasion identified by DCE methods. The first three columns appear in Table A7 and are shown here for comparison purposes. The fourth column implements the micro DCE2001 method used on 2001 NRP data by Johns and Slemrod (2010). We then report estimates using alternative macro allocation methods considered in Figure A14 in the final three columns. See Guyton et al. (2023) for additional details on DCE specifications.

Table A14: Comparing DCE Specifications: True Income as a % of Total True Income - TY2006-2013

Bin	Reported income (no sophisticated)	Exam-corrected income, no sophisticated	Exam-corrected income, add sophisticated	Baseline MA-DCE (Exam-corrected income share)	Micro DCE2001	MA-DCE Exam-detected under-rep share	MA-DCE Modified reported income share	MA-DCE Reported income share
P0-10	-2.6	-2.1	-2.0	-1.9	-0.1	-1.9	-1.6	-1.9
P10-20	1.0	1.0	1.0	1.0	0.3	1.0	1.0	1.0
P20-30	2.1	2.1	2.1	2.1	1.0	2.2	2.1	2.1
P30-40	3.2	3.4	3.3	3.3	1.6	3.3	3.4	3.2
P40-50	4.7	4.8	4.7	4.6	2.4	4.6	4.8	4.5
P50-60	6.4	6.5	6.4	6.3	3.8	6.2	6.4	6.1
P60-70	8.6	8.7	8.5	8.4	5.4	8.3	8.5	8.2
P70-80	11.7	11.6	11.4	11.3	8.4	11.1	11.4	11.1
P80-90	16.6	16.4	16.2	15.9	13.3	15.9	16.1	15.7
P90-95	12.0	11.8	11.6	11.6	12.9	11.5	11.7	11.4
P95-99	16.1	16.0	15.9	16.1	26.1	16.2	15.9	16.0
P99-99.5	4.3	4.2	4.3	4.4	7.5	4.4	4.2	4.5
P99.5-99.9	6.7	6.5	6.9	7.0	10.3	7.1	6.8	7.3
P99.9-P99.95	2.0	1.9	2.0	2.1	2.3	2.1	2.0	2.2
P99.95-P99.99	3.2	3.0	3.3	3.3	2.4	3.4	3.1	3.6
P99.99-100	4.2	4.1	4.5	4.6	2.4	4.7	4.2	4.9
Top 1%	20.3	19.8	21.0	21.4	24.8	21.7	20.4	22.5

Note: This table illustrates the sensitivity of the results in Table A8 to assumptions about the location in the income distribution of evasion identified by DCE methods. The first four columsn appear in Table A8 and are shown here for comparison purposes. The fifth column implements the micro DCE2001 method used on 2001 NRP data by Johns and Slemrod (2010). We then report estimates using alternative macro allocation methods considered in Figure A14 in the final three columns. See Guyton et al. (2023) for additional details on DCE specifications. Naturally, methods allocating less undetected evasion from DCE at the top imply smaller estimates of the top 1% income share. However, accounting for sophisticated evasion has a substantial impact regardless of the DCE allocation, and we observe that accounting for evasion, including sophisticated evasion, increases the top 1% share relative to reported incomes for any of the allocation methods we consider.

Table A15: Comparing Sophisticated Evasion Scenarios: Under-Reported Income as a % of Total Under-Reported Income, Using Micro DCE2019

	Exam-corrected (no sophisticated)	Benchmark sophisticated evasion	Lower bound sophisticated evasion	Upper bound sophisticated evasion	Aggregate sophisticated evasion on the low end	Aggregate sophisticated evasion on the high-end
P0-10	-0.7	0.7	1.0	0.7	0.7	0.7
P10-20	0.9	0.9	1.0	0.8	1.0	0.9
P20-30	1.7	1.9	2.0	1.6	1.9	1.8
P30-40	2.5	2.9	3.2	2.5	3.0	2.8
P40-50	3.1	4.6	5.0	4.0	4.8	4.4
P50-60	3.7	5.6	6.2	4.9	5.9	5.4
P60-70	5.5	7.4	8.1	6.5	7.7	7.1
P70-80	7.0	9.6	10.5	8.5	10.0	9.2
P80-90	9.8	13.1	14.3	11.5	13.5	12.5
P90-95	9.1	10.7	11.6	9.5	11.0	10.4
P95-99	12.0	18.1	18.9	16.7	18.2	17.9
P99-99.5	4.1	5.6	5.4	5.6	5.4	5.7
P99.5-99.9	4.6	8.3	7.0	9.4	7.8	9.0
P99.9-P99.95	0.9	2.1	1.3	2.9	1.9	2.5
P99.95-P99.99	1.5	4.0	2.6	5.5	3.5	4.6
P99.99-100	0.9	4.4	1.9	9.5	3.6	5.3
Top 1%	12.0	24.4	18.2	32.9	22.2	27.1

Note: This table examines the sensitivity of the results in Table 1 to assumptions about sophisticated evasion under the micro DCE2019 specification. We use the micro DCE2019 specification for DCE-identified under-reporting in all but the first column. The first two columns appear in Table Table 1. The remaining columns illustrate how the location of unreported income changes for the other scenarios for sophisticated evasion we considered. These are the same four alternatives to the benchmark scenarios depicted in Figure A15. Unsurprisingly, scenarios featuring evasion that is smaller and/or less concentrated at the top of the distribution reduce the concentration of unreported income, but accounting for sophisticated evasion remains a sizable modification to the distribution of unreported income in every scenario we consider.

TABLE A16: COMPARING SOPHISTICATED EVASION SCENARIOS: UNDER-REPORTED INCOME AS A % OF TOTAL UNDER-REPORTED INCOME, TY2008-2013, USING BASELINE MA-DCE

	Exam-corrected (no sophisticated)	Benchmark sophisticated evasion	Lower bound sophisticated evasion	Upper bound sophisticated evasion	Aggregate sophisticated evasion on the low end	Aggregate sophisticated evasion on the high-end
P0-10	-0.7	-0.1	0.1	0.0	-0.1	-0.1
P10-20	0.9	1.4	1.5	1.2	1.5	1.4
P20-30	1.7	2.5	2.8	2.2	2.6	2.4
P30-40	2.5	3.3	3.6	2.9	3.4	3.1
P40-50	3.1	4.0	4.4	3.5	4.2	3.8
P50-60	3.7	4.7	5.1	4.1	4.9	4.5
P60-70	5.5	6.5	7.1	5.7	6.7	6.2
P70-80	7.0	8.2	9.0	7.3	8.5	7.9
P80-90	9.8	11.3	12.3	10.0	11.7	10.9
P90-95	9.1	9.8	10.6	8.7	10.1	9.5
P95-99	12.0	16.3	17.0	15.2	16.4	16.2
P99-99.5	4.1	6.0	5.8	5.9	5.9	6.1
P99.5-99.9	4.6	9.6	8.4	10.5	9.1	10.2
P99.9-P99.95	0.9	3.0	2.3	3.6	2.8	3.3
P99.95-P99.99	1.5	5.3	4.0	6.6	4.9	5.8
P99.99-100	0.9	8.1	6.0	12.7	7.5	8.8
Top 1%	12.0	31.9	26.5	39.3	30.1	34.2

Note: This table examines the sensitivity of the results in Table 1 to assumptions about sophisticated evasion under the baseline macro allocation DCE specification. We use the baseline MA-DCE specification to allocate DCE-identified under-reporting in all but the first column. The first two columns appear in Table Table 1. The remaining columns illustrate how the location of unreported income changes for the other scenarios for sophisticated evasion we considered. These are the same four alternatives to the benchmark scenarios depicted in Figure A15. Unsurprisingly, scenarios featuring evasion that is smaller and/or less concentrated at the top of the distribution reduce the concentration of unreported income, but accounting for sophisticated evasion remains a sizable modification to the distribution of unreported income in every scenario we consider.

TABLE A17: COMPARING SOPHISTICATED EVASION SCENARIOS: UNDER-REPORTED INCOME AS A % OF TOTAL UNDER-REPORTED INCOME, TY2006-2013, USING BASELINE MA-DCE

	Exam-corrected (no sophisticated)	Benchmark sophisticated evasion	Lower bound sophisticated evasion	Upper bound sophisticated evasion	Aggregate sophisticated evasion on the low end	Aggregate sophisticated evasion on the high-end
P0-10	1.4	0.8	1.0	0.7	0.8	0.7
P10-20	0.7	0.7	0.8	0.6	0.7	0.7
P20-30	2.8	2.1	2.3	1.8	2.2	2.0
P30-40	4.9	3.2	3.5	2.8	3.4	3.1
P40-50	6.2	3.9	4.3	3.4	4.1	3.7
P50-60	7.6	4.9	5.3	4.3	5.1	4.7
P60-70	9.7	6.6	7.2	5.7	6.8	6.3
P70-80	11.9	8.6	9.4	7.5	8.9	8.2
P80-90	15.6	11.7	12.8	10.4	12.1	11.2
P90-95	12.2	9.8	10.6	8.7	10.1	9.5
P95-99	15.7	16.8	17.5	15.6	16.9	16.7
P99-99.5	3.9	5.9	5.7	5.8	5.7	6.0
P99.5-99.9	5.1	9.7	8.5	10.6	9.2	10.3
P99.9-P99.95	1.1	3.0	2.2	3.6	2.8	3.3
P99.95-P99.99	0.7	4.8	3.5	6.2	4.4	5.4
P99.99-100	0.6	7.5	5.3	12.2	6.9	8.2
Top 1%	11.4	30.9	25.3	38.4	29.0	33.2

Note: This table reports the sensitivity of the results in Table A7 to assumptions about sophisticated evasion. We use the baseline MA-DCE specification to allocate DCE-identified under-reporting in all but the first column. The first two columns appear in Table Table A7. The remaining columns illustrate how the location of unreported income changes for the other scenarios for sophisticated evasion we considered. These are the same four alternatives to the benchmark scenarios depicted in Figure A15. Unsurprisingly, scenarios featuring evasion that is smaller and/or less concentrated at the top of the distribution reduce the concentration of unreported income, but accounting for sophisticated evasion remains a sizable modification to the distribution of unreported income in every scenario we consider.

TABLE A18: COMPARING SOPHISTICATED EVASION SCENARIOS: TRUE INCOME AS A % OF TOTAL TRUE INCOME, TY2008-2013, USING MICRO DCE2019

	Reported income	Exam-corrected (no sophisticated)	Benchmark sophisticated evasion	Lower bound sophisticated evasion	Upper bound sophisticated evasion	Aggregate sophisticated evasion on the low end	Aggregate sophisticated evasion on the high-end
P0-10	-1.3	-1.0	-0.6	-0.6	-0.6	-0.6	-0.6
P10-20	1.4	1.5	1.5	1.5	1.5	1.5	1.5
P20-30	2.4	2.5	2.6	2.6	2.5	2.6	2.5
P30-40	3.5	3.6	3.7	3.7	3.6	3.7	3.6
P40-50	4.8	4.9	4.9	5.0	4.8	4.9	4.9
P50-60	6.4	6.5	6.4	6.5	6.3	6.5	6.4
P60-70	8.5	8.5	8.4	8.5	8.3	8.5	8.4
P70-80	11.5	11.5	11.2	11.4	11.0	11.3	11.2
P80-90	16.3	16.2	15.8	16.0	15.6	15.9	15.7
P90-95	11.8	11.7	11.4	11.5	11.3	11.5	11.4
P95-99	15.6	15.5	15.4	15.4	15.2	15.4	15.4
P99-99.5	4.0	4.0	4.0	4.0	4.0	4.0	4.0
P99.5-99.9	6.3	6.1	6.2	6.1	6.4	6.2	6.3
P99.9-P99.95	1.8	1.8	1.8	1.7	1.9	1.8	1.9
P99.95-P99.99	2.9	2.8	2.9	2.8	3.2	2.9	3.0
P99.99-100	4.1	4.0	4.1	3.9	4.8	4.0	4.2
Top 1%	19.1	18.7	19.2	18.4	20.4	18.9	19.5

Note: This table reports the sensitivity of the results in Table 2 to assumptions about sophisticated evasion. We employ the micro DCE2019 specification for DCE-identified under-reporting in all but the first two columns. The first three columns appear in Table 2. The remaining columns illustrate how the distribution of true income changes in the other scenarios for sophisticated evasion we considered. These are the same four alternatives to the benchmark scenarios depicted in Figure A15. Unsurprisingly, scenarios featuring evasion that is smaller and/or less concentrated at the top of the distribution reduce the concentration of income, but accounting for sophisticated evasion remains a significant modification to the distribution of true income in every scenario we consider.

TABLE A19: COMPARING SOPHISTICATED EVASION SCENARIOS: TRUE INCOME AS A % OF TOTAL TRUE INCOME, TY2008-2013, USING BASELINE MA-DCE

	Reported income	Exam-corrected (no sophisticated)	Benchmark sophisticated evasion	Lower bound sophisticated evasion	Upper bound sophisticated evasion	Aggregate sophisticated evasion on the low end	Aggregate sophisticated evasion on the high-end
P0-10	-1.3	-1.0	-0.8	-0.8	-0.8	-0.9	-0.8
P10-20	1.4	1.5	1.5	1.5	1.4	1.5	1.5
P20-30	2.4	2.5	2.5	2.5	2.4	2.5	2.5
P30-40	3.5	3.6	3.5	3.6	3.5	3.6	3.5
P40-50	4.8	4.9	4.8	4.8	4.7	4.8	4.7
P50-60	6.4	6.5	6.2	6.3	6.1	6.3	6.2
P60-70	8.5	8.5	8.3	8.3	8.1	8.3	8.2
P70-80	11.5	11.5	11.1	11.2	10.9	11.1	11.0
P80-90	16.3	16.2	15.6	15.8	15.4	15.7	15.6
P90-95	11.8	11.7	11.4	11.5	11.2	11.4	11.3
P95-99	15.6	15.5	15.6	15.7	15.4	15.6	15.6
P99-99.5	4.0	4.0	4.2	4.1	4.2	4.1	4.2
P99.5-99.9	6.3	6.1	6.6	6.4	6.8	6.5	6.7
P99.9-P99.95	1.8	1.8	1.9	1.9	2.1	1.9	2.0
P99.95-P99.99	2.9	2.8	3.2	3.0	3.4	3.1	3.2
P99.99-100	4.1	4.0	4.6	4.3	5.3	4.5	4.7
Top 1%	19.1	18.7	20.4	19.7	21.7	20.2	20.8

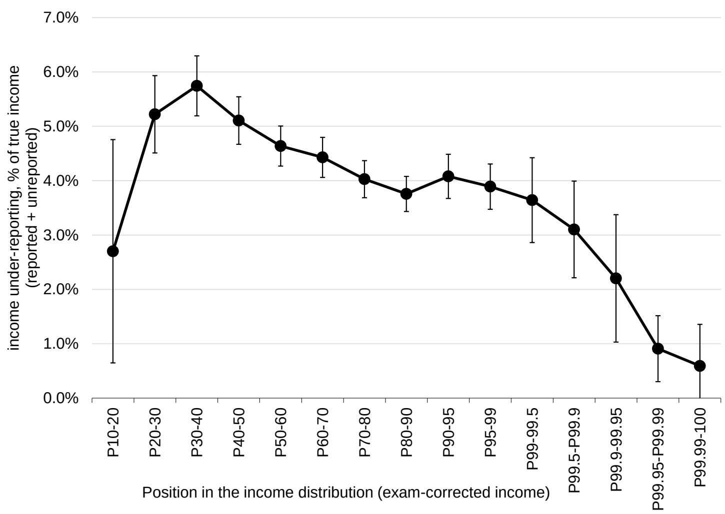
Note: This table reports the sensitivity of the results in Table 2 to assumptions about sophisticated evasion. We use the baseline MA-DCE specification to allocate DCE-identified under-reporting in all but the first two columns. The first three columns appear in Table 2. The remaining columns illustrate how the distribution of true income changes in the other scenarios for sophisticated evasion we considered. These are the same four alternatives to the benchmark scenarios depicted in Figure A15. Unsurprisingly, scenarios featuring evasion that is smaller and/or less concentrated at the top of the distribution reduce the concentration of income, but accounting for sophisticated evasion remains a significant modification to the distribution of true income in every scenario we consider.

TABLE A20: COMPARING SOPHISTICATED EVASION SCENARIOS: TRUE INCOME AS A % OF TOTAL TRUE INCOME, TY2006-2013, USING BASELINE MA-DCE

	Reported income	Exam-corrected (no sophisticated)	Benchmark sophisticated evasion	Lower bound sophisticated evasion	Upper bound sophisticated evasion	Aggregate sophisticated evasion on the low end	Aggregate sophisticated evasion on the high-end
P0-10	-2.6	-2.1	-1.9	-1.9	-1.8	-1.9	-1.9
P10-20	1.0	1.0	1.0	1.0	1.0	1.0	1.0
P20-30	2.1	2.1	2.1	2.1	2.1	2.1	2.1
P30-40	3.2	3.4	3.3	3.3	3.2	3.3	3.3
P40-50	4.7	4.8	4.6	4.7	4.6	4.7	4.6
P50-60	6.4	6.5	6.3	6.3	6.1	6.3	6.2
P60-70	8.6	8.7	8.4	8.5	8.2	8.4	8.3
P70-80	11.7	11.6	11.3	11.4	11.1	11.3	11.2
P80-90	16.6	16.4	15.9	16.1	15.7	16.0	15.8
P90-95	12.0	11.8	11.6	11.7	11.4	11.6	11.5
P95-99	16.1	16.0	16.1	16.1	15.9	16.1	16.0
P99-99.5	4.3	4.2	4.4	4.4	4.4	4.4	4.4
P99.5-99.9	6.7	6.5	7.0	6.8	7.1	6.9	7.1
P99.9-P99.95	2.0	1.9	2.1	2.0	2.2	2.0	2.1
P99.95-P99.99	3.2	3.0	3.3	3.2	3.5	3.3	3.4
P99.99-100	4.2	4.1	4.6	4.4	5.3	4.5	4.7
Top 1%	20.3	19.8	21.4	20.7	22.6	21.1	21.8

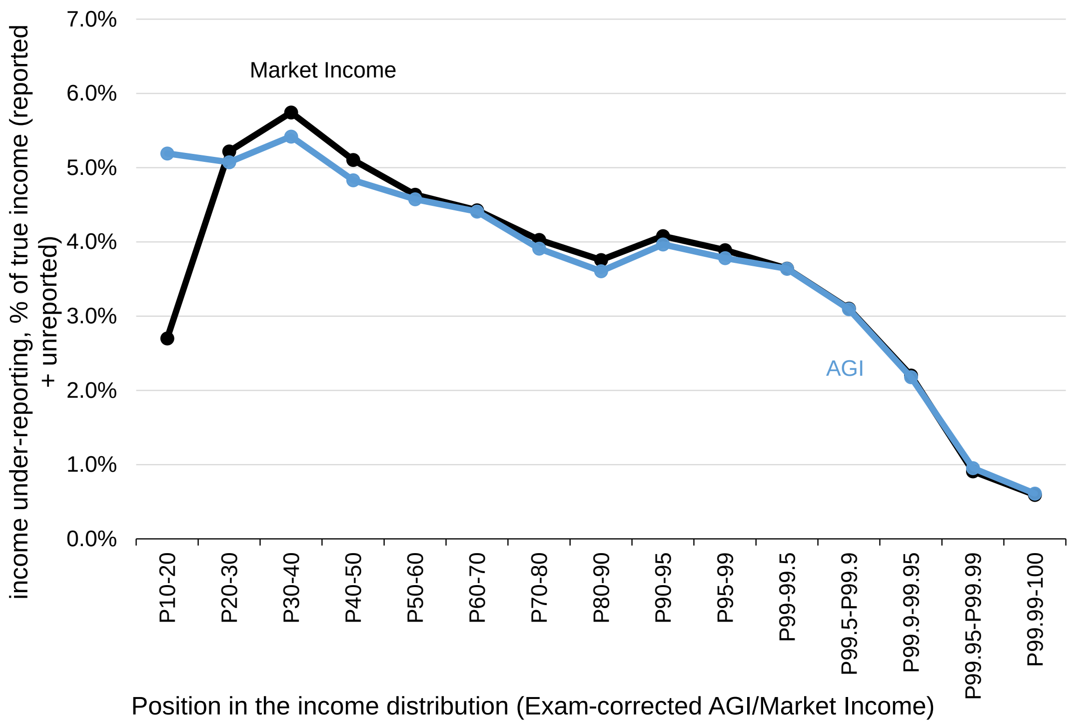
Note: This table reports the sensitivity of the results in Table A8 to assumptions about sophisticated evasion. We use the baseline MA-DCE specification to allocate DCE-identified under-reporting in all but the first two columns. The first three columns appear in Table A8. The remaining columns illustrate how the distribution of true income changes in the other scenarios for sophisticated evasion we considered. These are the same four alternatives to the benchmark scenarios depicted in Figure A15. Unsurprisingly, scenarios featuring evasion that is smaller and/or less concentrated at the top of the distribution reduce the concentration of income, but accounting for sophisticated evasion remains a significant modification to the distribution of true income in every scenario we consider.

FIGURE A1: UNREPORTED INCOME DETECTED IN RANDOM AUDITS: ESTIMATES AND 95% CONFIDENCE INTERVALS



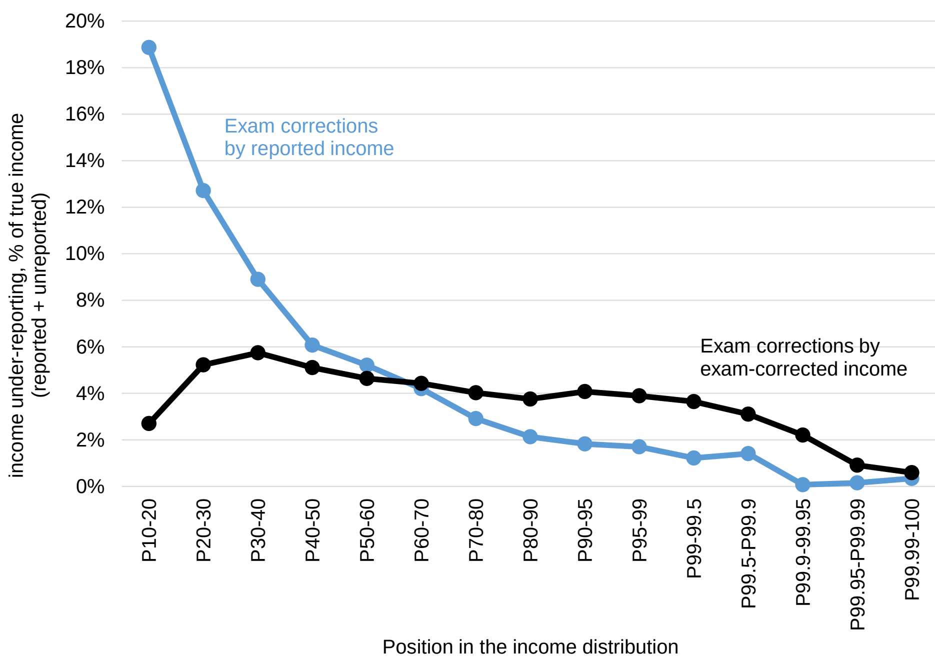
This figure reports the main estimate and 95% confidence interval for the estimates in Figure 1a of the main text. We observe that the profile of evasion depicted in Figure 1a is relatively precisely estimated.

FIGURE A2: SENSITIVITY ANALYSIS FOR THE DEFINITION OF INCOME



Note: This figure depicts the sensitivity of the results in Figure 1a to the definition of income we use. Market income is total income reported on form 1040 minus Social Security benefits, unemployment insurance benefits, alimony, and state refunds. We observe that the use of market income versus AGI is immaterial except for the bottom of the distribution (e.g., because non-compliance on Social Security benefits is disregarded when using market income).

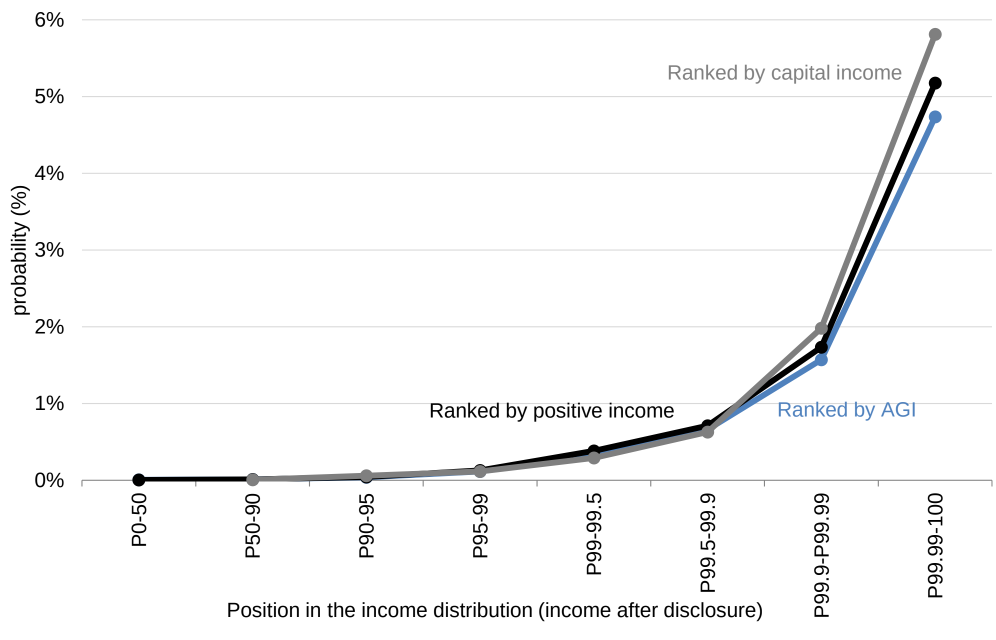
Figure A3: The Influence of Re-Ranking on Estimated Rates of Income Under-Reporting



Note: This figure illustrates the impact of re-ranking on the profile of income under-reporting through the income distribution using NRP random audit data. We begin with "Exam corrections by reported income," which ranks taxpayers by originally reported income and calculates income and under-reporting gaps using exam corrections only (i.e., not including undetected under-reporting identified by DCE). We then continue to use exam corrections only but re-rank individuals by exam-corrected income in "exam corrections by exam-corrected income," which matches Figure 1a. We find that this re-ranking substantially decreases estimated rates of evasion in the bottom 50% of the distribution.

FIGURE A4: PROBABILITY OF FILING AN FBAR FOR THE FIRST TIME IN 2009-11

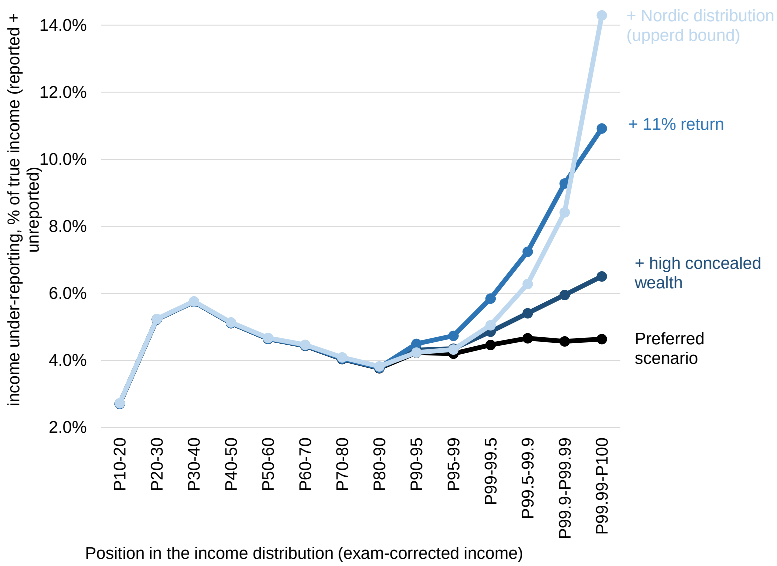
(Haven accounts only, U.S. filers only)



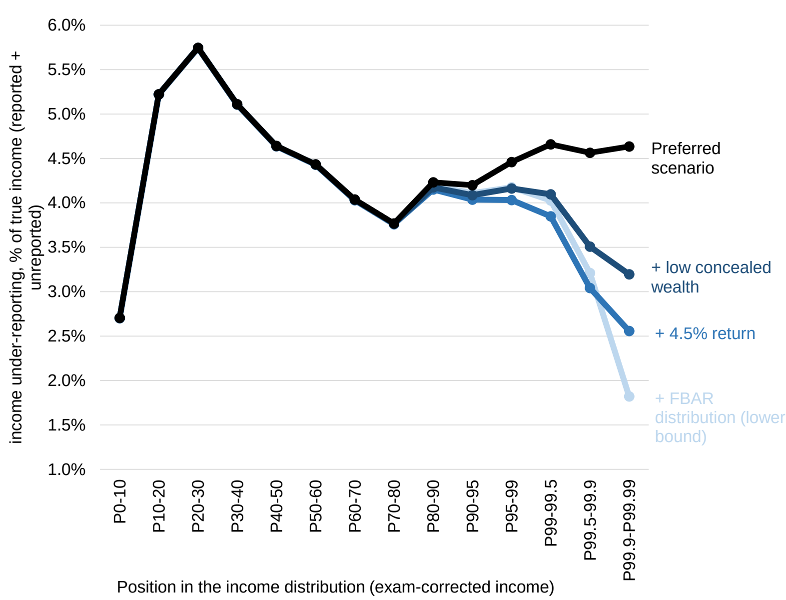
Note: This figure plots the fraction of the population within each part of the income distribution that are present in the first-time FBAR filer sample. We observe that the probability of being in the sample is much higher at the very top of the income distribution, with a nearly trivial fraction of the bottom 99 percent of the income distribution disclosing an offshore account. We observe that the overall profile is very similar for the three different income concepts, though it is steepest for capital income, followed by positive income.

Figure A5: Sensitivity Analysis for Offshore Wealth: Decomposition

(a) Upper Bound



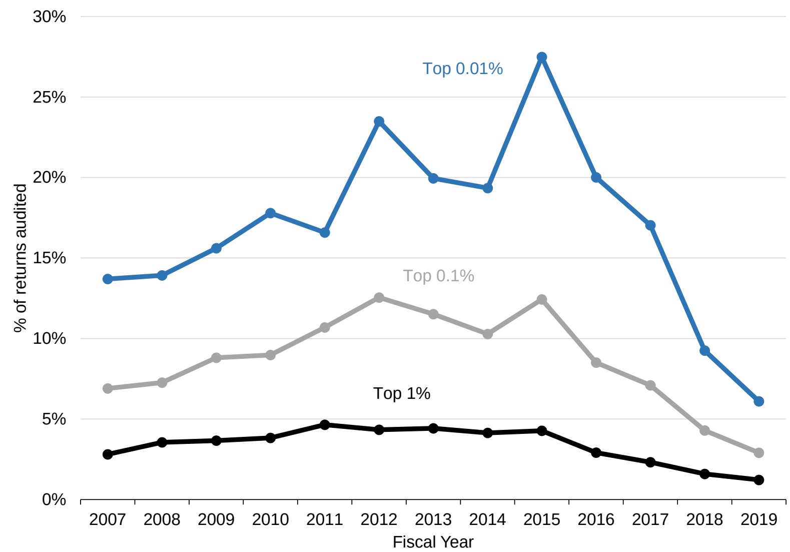
(b) Lower Bound



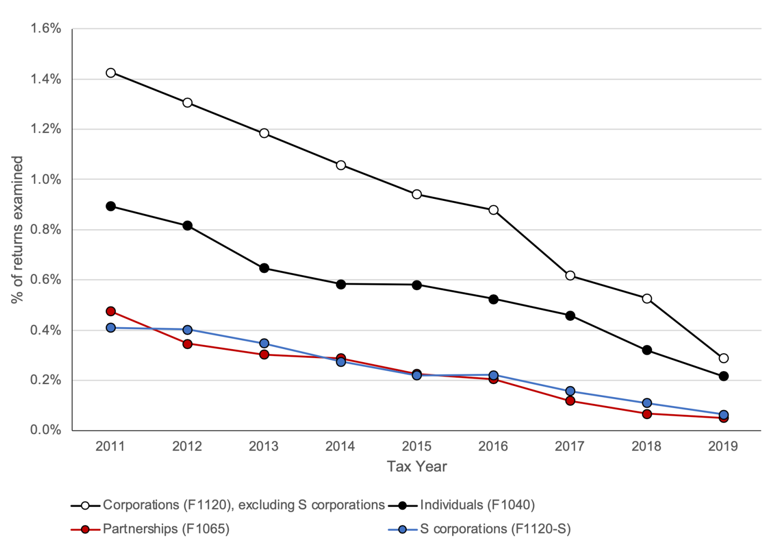
Note: This figure plots unreported income over total true income by rank in the income distribution with and without accounting for offshore wealth. We illustrate how different assumptions contribute to the upper and lower bounds illustrated in Figure 4b. In either figure, we begin with our preferred scenario for offshore wealth and then progressively add assumptions for the alternative scenarios described in Table A3.

FIGURE A6: AUDIT RATES OVER TIME

(a) High-income individual returns



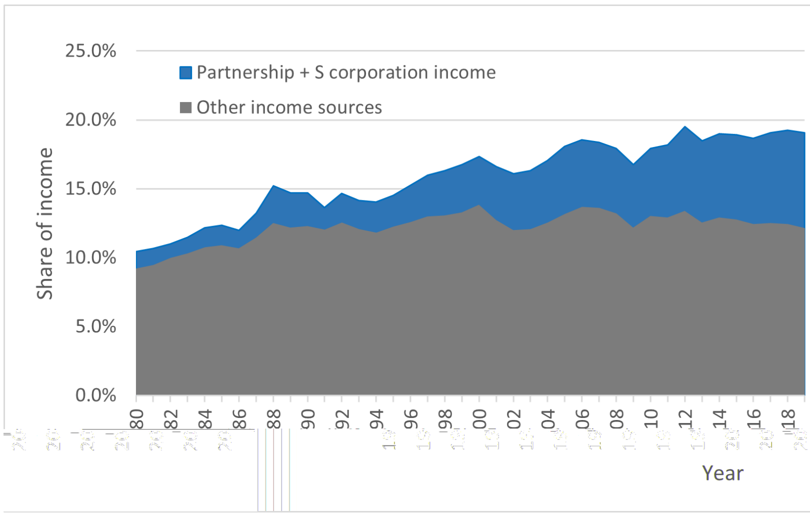
(b) Individual versus entity returns



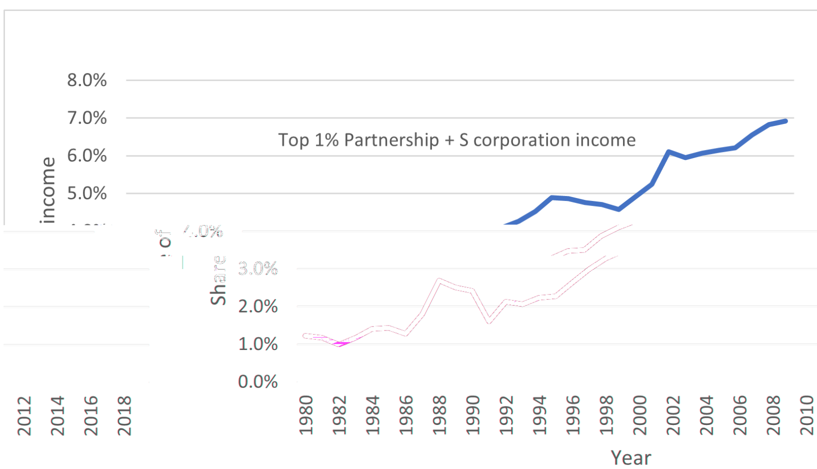
Note: The first panel of this figure plots the share of individual tax returns that are subject to audit over time in the three income groups. We observe that audit rates are highest at the very top, and they increase and then decline through the period of observation period. The second panel plots audit rates over time for individuals, partnerships, S corporations, and all other types of corporations. We observe that audit rates of all these types of returns have fallen substantially since 2011. Audit rate for partnerships and S corporations were already relatively low in 2011, and they decreaed to less than 0.1% for 2019. The first panel estimates audit rates using operational audit micro data, in which we observe the fiscal year in which the return was originally filed, while the second panel plots publicly available audit rates by tax year from the IRS Databook IRS (2022), Table 17.

FIGURE A7: IMPORTANCE OF PASS-THROUGH INCOME AT THE TOP OVER TIME

(a) Share of income to the top 1% by source



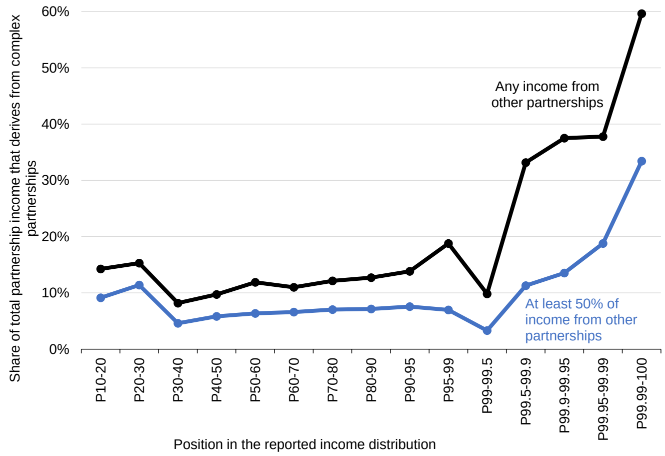
(b) Top 1% partnership and S corporation income as a share of total income



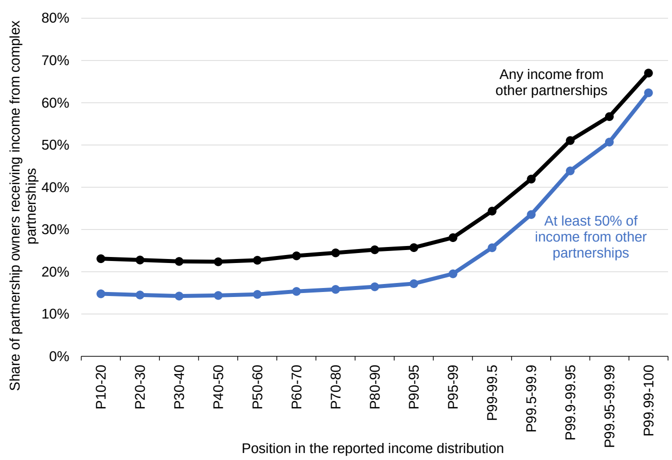
Note: The first panel shows the share of total income going to the top 1% of the income distribution from 1980-2019, splitting income from partnerships and S corporations from all other sources. Top 1% income shares haven risen substantially over time (by 1.83 times), and the majority of that growth is attributable to pass-through income, which rises from 11 to 36% of top 1% income over this period. The second panel isolates the trend in partnership and S corporation income going to the top 1% as a share of total income, which rises steadily from 1% to 7% over this period. Income shares are pre-tax income from the Distributional National Accounts (DINA) tables from Piketty et al. (2018).

FIGURE A8: THE IMPORTANCE OF COMPLEX (TIERED) PARTNERSHIP STRUCTURES AT THE TOP OF THE DISTRIBUTION

(a) Share of total partnership income that derives from complex partnerships



(b) Share of partnership owners receiving income from complex partnerships



Note: This figured illustrates how partnership complexity evolves through the income distribution. Generally, we consider that an individual has an interest in a complex partnership if they receive income from a partnership that in turn receives income from other partnerships, i.e. if the individual has an interest in a tiered partnership. We operationalize this definition in two ways: a complex partnership is defined as one receiving 1) *any* income from other partnerships, or 2) at least half of its income from other partnerships. In panel a) we plot the share of all partnership income in a given (reported) income bin that derives from complex partnerships. In panel b), we plot the share of individual partnership owners who receive income from a complex partnership in each income bin. For each of these, we observe that complex partnerships grow rapidly in importance within the top 1% of the income distribution.

FIGURE A9: ENTITY-LEVEL PASS-THROUGH UNDER-REPORTING: SENSITIVITY ANALYSIS AROUND PASS-THROUGH BUSINESS INCOME

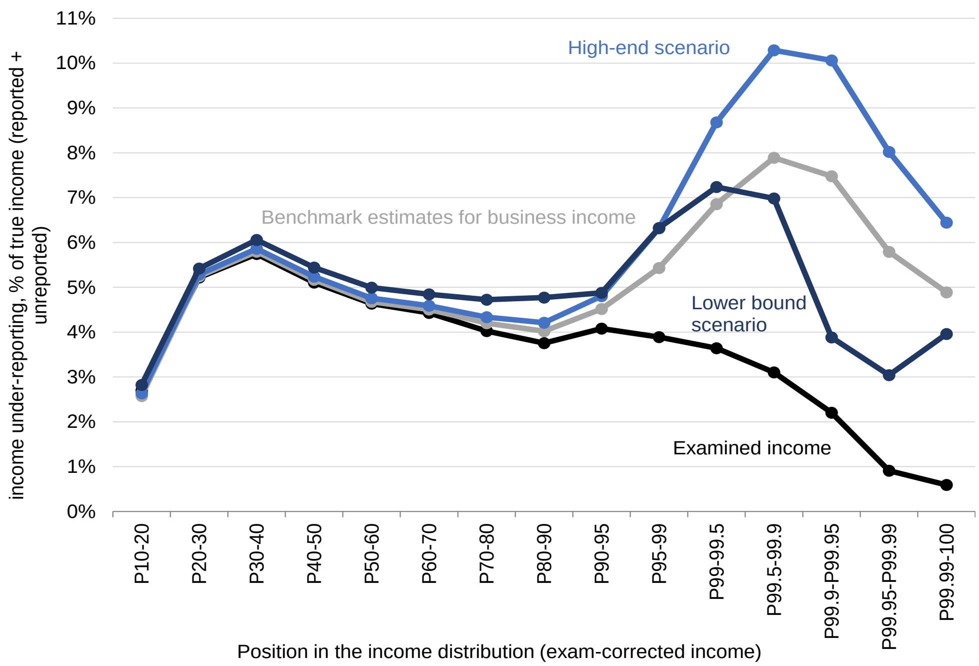


FIGURE A10: ENTITY-LEVEL PASS-THROUGH UNDER-REPORTING: SENSITIVITY ANALYSIS AROUND PASS-THROUGH INVESTMENT INCOME

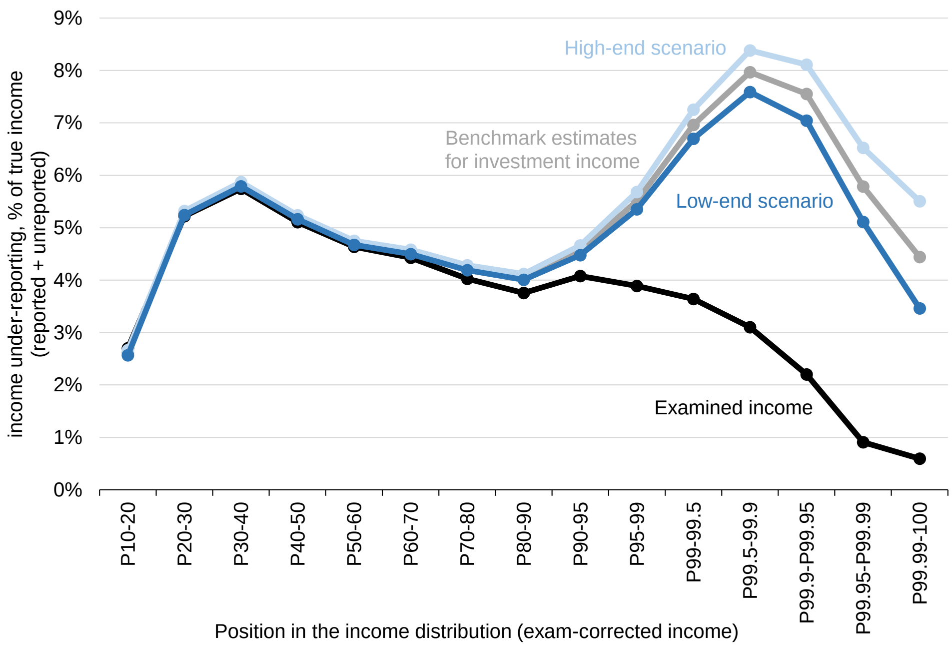


FIGURE A11: ENTITY-LEVEL PASS-THROUGH UNDER-REPORTING: THE EFFECT OF ACCOUNTING FOR PASS-THROUGH BUSINESS LOSSES

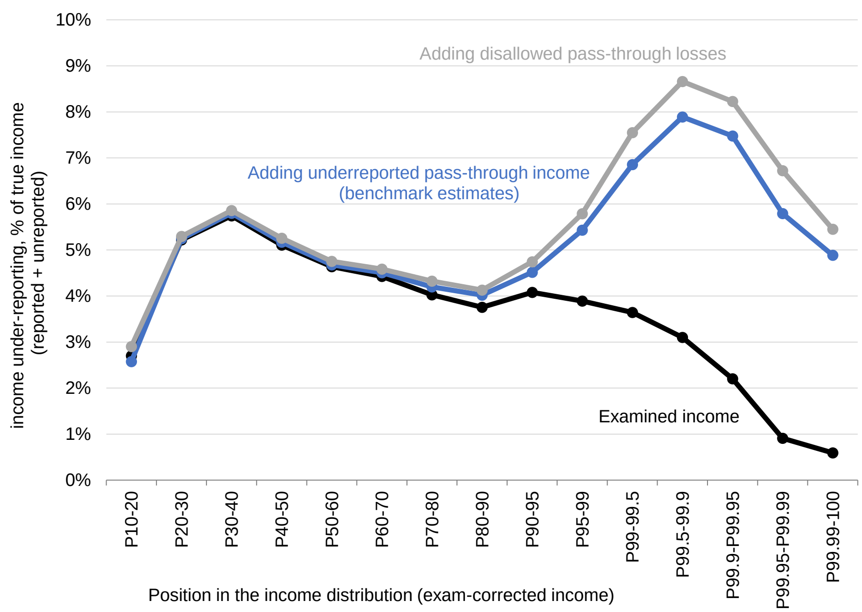


FIGURE A12: ENTITY-LEVEL PASS-THROUGH UNDER-REPORTING: THE EFFECT OF ACCOUNTING FOR CIRCULAR PARTNERSHIPS

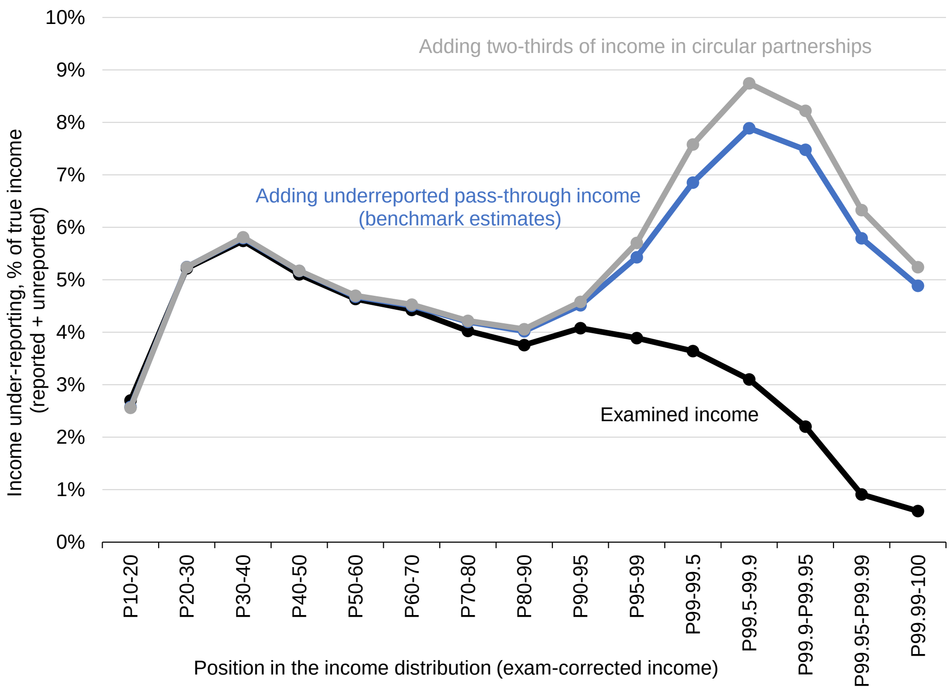
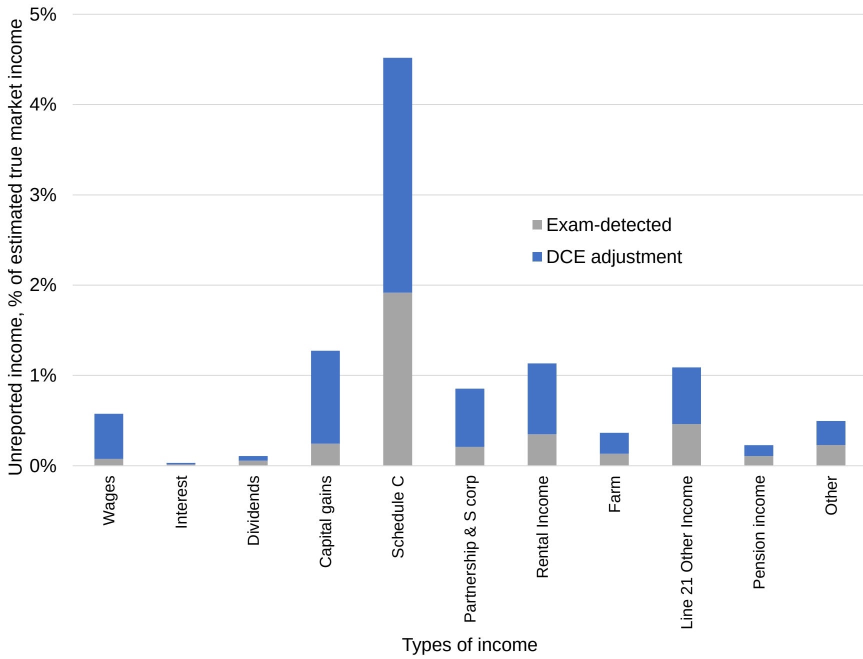
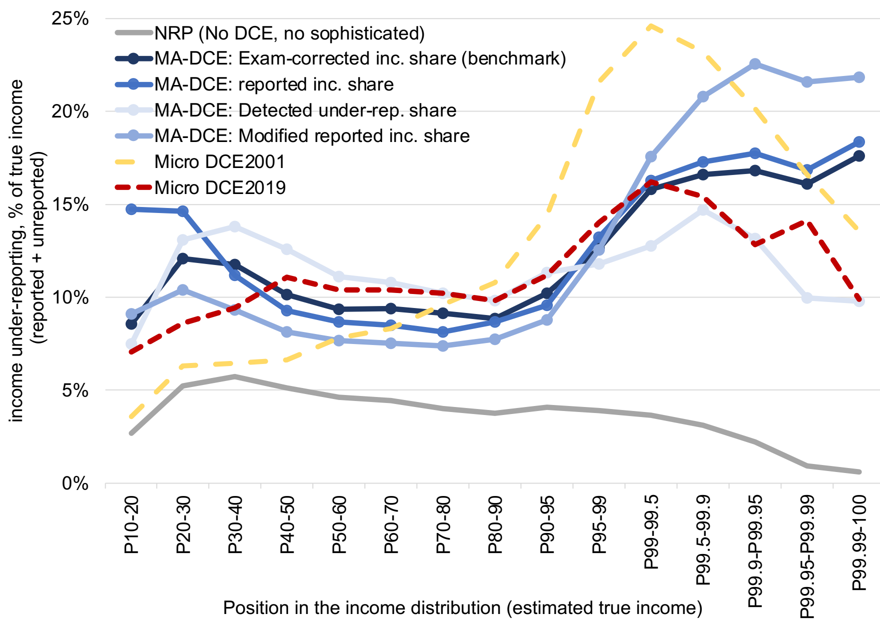


FIGURE A13: DETECTION-CONTROLLED ESTIMATION: TOTAL UNDER-REPORTED INCOME BY TYPE OF INCOME, % OF ESTIMATED TRUE MARKET INCOME



Note: This figure illustrates estimated detected and undetected under-reporting in DCE2019 by type of income. We scale both exam-detected and DCE adjusted under-reporting totals by estimated total true income (= reported income + detected under-reporting + undetected under-reporting). Note also that what IRS (2019) call "Form 1040 Other Income" is referred to as "Line 21 Other Income" here, as this item appears on Line 21 of the Form 1040, while the residual "Other" category in the last bar of the figure refers to all other components of income.

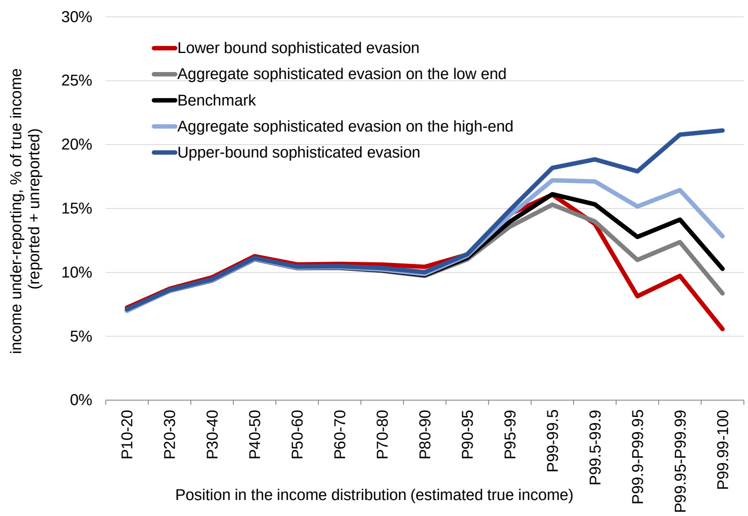
FIGURE A14: SENSITIVITY ANALYSIS FOR ALTERNATIVE DCE SPECIFICATIONS: UNDER-REPORTED INCOME (% TRUE INCOME)



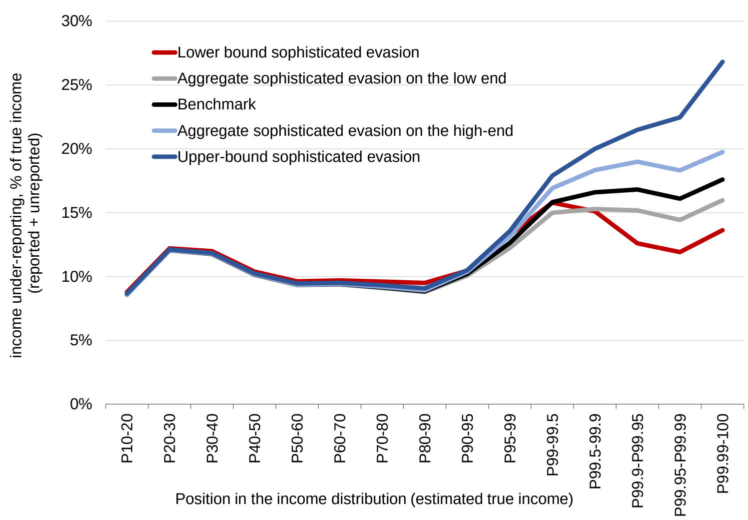
Note: This figure plots the our main estimates of total under-reporting as a fraction of total true income, including sophisticated evasion and DCE, for the different scenarios for the allocation of undetected under-reporting identified by DCE methods, as described in Section 5 and Guyton et al. (2023). The difference between this figure and Figure 8c derive entirely from the inclusion of sophisticated evasion here.

FIGURE A15: SENSITIVITY ANALYSIS FOR ASSUMPTIONS ABOUT SOPHISTICATED EVASION: UNREPORTED INCOME (% OF TRUE INCOME)

(a) Micro DCE2019



(b) Baseline MA-DCE



Note: This figure plots the our main estimates of total under-reporting as a fraction of total true income, for the different scenarios for sophisticated evasion described in Section 6.1.

B Proofs

Lemma 2. *Under Assumption 1, as y becomes arbitrarily large, $g_1(y, p_1) - g_0(y, p_1)$ converges to zero.*

Proof. We can re-express the optimization problem with (g, a) as the choice variables rather than (e, a). We denote the fixed cost κ as a share of income by $\tilde{\kappa} = k/y$. We can then express consumption in the detected and undetected state by $c_D = (1 - \tau - \theta\tau g - \tilde{\kappa}a)y$, and $c_N = (1 - \tau + \tau g - \tilde{\kappa}a)y$. respectively. The first order condition with respect to g of the optimization problem in equation 1 is

$$\frac{u'(c_D)}{u'(c_N)} = \frac{1-p}{p\theta}. \tag{3}$$

We wish to compare $g_1(y, p_1)$ and $g_0(y, p_1)$ at large y. As both of these are evaluated at $p = p_1$, the right-hand side of (3) is constant for this comparison. Comparing the first order conditions under $a = 1$ and $a = 0$, we have:

$$\frac{u'((1 - \tau - \theta\tau g_1(y, p_1) - \tilde{\kappa})y)}{u'((1 - \tau + \tau g_1(y, p_1) - \tilde{\kappa})y)} = \frac{u'((1 - \tau - \theta\tau g_0(y, p_1))y)}{u'((1 - \tau + \tau g_0(y, p_1))y)}. \tag{4}$$

As y becomes large, $\tilde{\kappa} = \kappa/y$ becomes arbitrarily small. As $u'' < 0$, the LHS and RHS of equation (4) are invertible in e. Finally, by Assumption 1, for arbitrarily large y, $g_0(y, p_1)$ on the RHS converges to a strictly positive constant. Altogether, it follows that for sufficiently large y, the $\tilde{\kappa}$ term on the LHS can be made arbitrarily small. The LHS can thus be made arbitrarily close to the RHS, so that $g_1(y, p_1)$ becomes arbitrarily close to $g_0(y, p_1)$. Equation 4 gives strong intuition about the validity of the lemma. We provide below a formalized proof of the convergence in the ϵ-δ sense.

We change the arguments of g from how they are defined above, as $p_1 = p$ is constant across g_1 and g_0 and the only element that makes the two different is the presence of $\tilde{\kappa}$. We define the forms that we use here as follows (note that $g(y, \tilde{\kappa}(y)) = g_1(y, p_1)$ and $g(y, 0) = g_0(y, p_1)$ in comparison with the MRS equation).

$$g(y, \tilde{\kappa}(y)) = \text{argmax}_{g\in[0,1]}(1 - p)u((1 - \tau + \tau g - \tilde{\kappa})y) + pu((1 - \tau - \tau\theta g - \tilde{\kappa})y) \tag{5}$$

$$g(y, 0) = \text{argmax}_{g\in[0,1]}(1 - p)u((1 - \tau + \tau g)y) + pu((1 - \tau - \tau\theta g)y) \tag{6}$$

As κ is a constant, we know that : $lim_{y->\infty}\tilde{\kappa}(y) = 0$.
Then, by the definition of limits, we know that for any $\delta > 0$, there exists a $c \in \mathbf{R}$ such that :

$$y > c \quad \Rightarrow \quad |\tilde{\kappa}(y) - 0| < \delta \tag{7}$$

Set $\epsilon > 0$, by continuity of g on real positive numbers, there exists some $c \in \mathbf{R}$,

$$y > c \quad \Rightarrow \quad |\tilde{\kappa}(y) - 0| < \delta \quad \Rightarrow \quad |g(y, \tilde{\kappa}(y)) - g(y, 0)| < \epsilon \tag{8}$$

Then, $g(y, \tilde{\kappa})$ converges to $g(y, 0)$ as y becomes arbitrarily large. Assumption 1 ensures that, as y becomes arbitrarily large, $g(y, 0)$ will be arbitrarily close to its non-zero limit.

□

Proposition 1. ***High-Income Concealment.*** *Under Assumption 1, there is a cutoff in the model $\hat{y}$ such that holding all else fixed, $y > \hat{y} \implies a = 1$ is optimal.*

Proof. We want to show that for a sufficiently large y, the difference in expected utility between $a = 1$ and $a = 0$ given optimal g_1 and g_0 must be positive. We express expected utility as a function of a and g_a as

$$U(p, \kappa, y) = (1 - p)u((1 - \tau + \tau g(p, \kappa, y) - \tilde{\kappa})y) + pu((1 - \tau - \tau\theta g(p, y, \kappa) - \tilde{\kappa})y), \tag{9}$$

where $g(p, \kappa, y)$ denotes the optimal level of evasion as a fraction of income, e/y, given the primitives. The difference between utility under adoption and non-adoption, given optimal evasion, is simply

$$\Delta_a U = U(p_1, \kappa, y) - U(p_0, 0, y). \tag{10}$$

The key to making use of Lemma 2 is to benchmark these expected utilities to expected utility under $\kappa = 0$ and $p = p_1$ - in which case behavior is given by $g(p_1, y, 0) = g_0(p_1, y)$, and expected utility by $U(p_1, 0, y)$. Adding and subtracting this from both sides of the above expression, we obtain:

$$\Delta_a U = [U(p_1, \kappa, y) - U(p_1, 0, y)] + \{U(p_1, 0, y) - U(p_0, 0, y)\} \tag{11}$$

Equation (11) decomposes $\Delta_a U$ into the difference due to the incursion of the cost - the first term in square brackets - and the difference due to the lower probability of detection - the second term, in curly brackets. The remaining structure of the proof shows that under Assumption 1, the latter dominates the former for large y.

Using the second fundamental theorem of calculus, we can rewrite the term in curly brackets above as

$$U(p_1, 0, y) - U(p_0, 0, y) = -\int_{p_1}^{p_0} U_p(p, 0, y) dp, \tag{12}$$

where $U_p(p, 0, y)$ is the partial derivative of U with respect to p evaluated at $(p, 0, y)$. Using the envelope theorem to characterize $U_p(p,0,y)$, we have

$$U(p_1,0,y)-U(p_0,0,y)=\int_{p_1}^{p_0}[u((1-\tau+\tau g(p,0,y))y)-u((1-\tau-\tau\theta g(p,0,y))y)]dp. \tag{13}$$

Note that provided $g(p,0,y)\neq 0$ for $p\in[p_1,p_0]$, this expression is strictly positive, because $p_1<p_0$ and $u'>0$. In words, provided the individual actually does evade some tax, decreasing the detection probability strictly increases expected utility.

To simplify expressions, as in the proof of Lemma 2, we define the argument of the utility function in the detected and undetected state given behavior $g(p,\kappa,y)$ by $c_D(p,\kappa,y)$ and $c_N(p,\kappa,y)$ respectively. Using equation (13) and the definition of U, we can rewrite equation (11) as

$$\begin{aligned}\Delta_a U =&(1-p_1)[u(c_N(p_1,\kappa,y))-u(c_N(p_1,0,y))]+p_1[u(c_D(p_1,\kappa,y))-u(c_D(p_1,0,y))]\\ &+\int_{p_1}^{p_0}[u(c_N(p,0,y))-u(c_D(p,0,y))]dp.\end{aligned} \tag{14}$$

Next, we use the second fundamental theorem of calculus again to express all the differences in utilities in the above equation as integrals of marginal utility over the appropriate range of final consumption. To understand these integrals, it helps to note that both $c_N(p,\kappa,y)$ and $c_N(p,\kappa,y)$ are decreasing in κ.[69] We write all integrals so that the lower limit of integration is less than the upper limit.

$$\Delta_a U=-(1-p_1)\int_{c_N(p_1,\kappa,y)}^{c_N(p_1,0,y)}u'(c)dc-p_1\int_{c_D(p_1,\kappa,y)}^{c_D(p_1,0,y)}u'(c)dc+\int_{p_1}^{p_0}\int_{c_D(p,0,y)}^{c_N(p,0,y)}u'(c)dcdp. \tag{15}$$

We now use diminishing marginal utility to find a simpler function $f(y)$ such that $\Delta_a U>f(y)$ always, and then construct an argument that $f(y)>0$ for sufficiently large values of y. For integrals with a positive sign in front (the third term), we construct f so that the integral is evaluated as a constant at the smallest u' over the specified range, which by $u''<0$ corresponds to u' at the upper limit of integration. For integrals with a negative sign in front (the first two terms), we should use the lower limit of integration. We thereby obtain

$$\begin{aligned}\Delta_a U >&-(1-p_1)[c_N(p_1,0,y)-c_N(p_1,\kappa,y)]u'(c_N(p_1,\kappa,y))\\ &-p_1[c_D(p_1,0,y)-c_D(p_1,\kappa,y)]u'(c_D(p_1,\kappa,y))\\ &+\int_{p_1}^{p_0}[c_N(p,0,y)-c_D(p,0,y)]u'(c_N(p,0,y))dp.\end{aligned} \tag{16}$$

[69] Differentiating the first-order condition in equation 3, we have $u''(c_N)\frac{\partial c_N}{\partial\kappa}=u''(c_D)\frac{\partial c_D}{\partial\kappa}$. This implies that the sign of $\frac{\partial c_D}{\partial\kappa}$ and $\frac{\partial c_D}{\partial\kappa}$. These two cannot both be positive, because this would imply that evasion is both increasing in κ (from $\frac{\partial c_N}{\partial\kappa}>0$) and decreasing (from $\frac{\partial c_D}{\partial\kappa}>0$). Hence they are both negative.

We modify $f(y)$ slightly by noting that from the first-order condition in equation 3,

$$u'(c_D(p_1, \kappa, y) = u'(c_N(p_1, \kappa, y)\frac{1-p_1}{\theta p_1}.$$

We also note that we can shrink the expression further by evaluating the last term with a constant marginal utility $u'(c_N(p_1, 0, y))$, as c_N is decreasing in p and $u'' < 0$. Substituting this into equation (16) and simplifying, we obtain

$$\begin{aligned}\Delta_a U > &-(1-p_1)u'(c_N(p_1,\kappa,y))\left\{c_N(p_1,0,y) - c_N(p_1,\kappa,y) + \theta^{-1}[c_D(p_1,0,y) - c_D(p_1,\kappa,y)])\right\}\\ &+ u'(c_N(p_1,0,y))\int_{p_1}^{p_0}[c_N(p,0,y) - c_D(p,0,y)]dp. \end{aligned} \tag{17}$$

We note that by construction $c_N(p,0,y) - c_D(p,0,y) = \tau(1+\theta)g(p,0,y)y$. As this expression is decreasing in p by Lemma 1, we shrink the function by evaluating it at the upper limit of integration. In so doing we arrive at an $f(y)$ that is simple enough to analyze for large y.:

$$\begin{aligned}\Delta_a U > f(y) \equiv &-(1-p_1)u'(c_N(p_1,\kappa,y))\left\{c_N(p_1,0,y) - c_N(p_1,\kappa,y) + \theta^{-1}[c_D(p_1,0,y) - c_D(p_1,\kappa,y)])\right\}\\ &+ u'(c_N(p_1,0,y))(p_0-p_1)\tau(1+\theta)g(p_0,0,y)y. \end{aligned} \tag{18}$$

As $u' > 0$ we find that[70]

$$\begin{aligned}f(y) > 0 \iff &-(1-p_1)\left\{c_N(p_1,0,y) - c_N(p_1,\kappa,y) + \theta^{-1}[c_D(p_1,0,y) - c_D(p_1,\kappa,y)])\right\}\\ &+ \frac{u'(c_N(p_1,0,y)}{u'(c_N(p_1,\kappa,y))}(p_0-p_1)\tau(1+\theta)g(p_0,0,y)y > 0 \end{aligned} \tag{19}$$

We now examine the behavior of the expression in equation (19) at large y. We know from Lemma 2 that $c_N(p_1,0,y) - c_N(p_1,\kappa,y)$ and $c_D(p_1,0,y) - c_D(p_1,\kappa,y)$ both become arbitrarily small as y becomes large. The term in the top row can therefore be made arbitrarily small. From Lemma 2, we also know that $\frac{u^{\theta}(c_N(p_1,0,y))}{u^{\theta}(c_N(p_1,\kappa,y)}$ converges to unity as y becomes large. Assumption 1 ensures that the second part of the term in the bottom row, $\tau(1+\theta)g(p,0,y)y$ grows arbitrarily large for large y. It follows that $f(y) > 0$ for sufficiently large y, and thus that $\Delta_a U > 0$ for sufficiently large y. □

Proposition 2. ***Incentivizing Concealment.*** *Suppose a policy increases the probability of detection only if $a = 0$. This policy will increase concealment.*

Proof. This result follows immediately from the envelope theorem. Differentiating $\Delta_a U$ with re-

[70]If u' converges to a strictly positive constant for arbitrarily large y, the proof from this point is more straightforward than what we present here. The result essentially follows directly from Assumption 1 and Lemma 2, which guarantee that the term in the top row shrinks while the term in the bottom row grows large. We construct the proof the way that we do to handle the case where u' approaches zero for large y, which is widely considered to be relevant.

spect to p_0 and applying the envelope theorem, we obtain

$$\frac{\partial \Delta_a U}{\partial p_0} = u(c_N(p_0, 0, y)) - u(c_D(p_0, 0, y)) > 0. \tag{20}$$

$\square$

Proposition 3. ***Comparative Statics of the Resource-Constrained Model.*** *In the optimization problem described by equation* (23)*,*

- $\frac{\partial N_h}{\partial c_h} < 0$
- $\frac{\partial N_l}{\partial c_h} > 0$ if and only if $-N_h R_h''/R_h' < 1$.

Proof. We solve the resource constraint for N_l in equation(23) and substitute this into the right-hand side of 25. We differentiate the resulting expression with respect to c_h and solve for $\frac{\partial N_h}{\partial c_h}$ to obtain:

$$\frac{\partial N_h}{\partial c_h} = \frac{R_h'(N_h) - N_h R_l''(N_l)}{c_h R''(N_h) - R'(N_h)} < 0 \tag{21}$$

The first result then follows from $R_\theta' > 0$ and $R_\theta'' < 0$ for each type $\theta = 0, 1$.

Proceeding similarly for N_l, we obtain

$$\frac{-(R_h'' N_h + R_h')c_l}{c_l^2 R_h'' + c_h^2 R_l''}. \tag{22}$$

This expression is positive whenever $N_h R_h'' + R_h' > 0 \iff -N_h R_h''/R_h' > 1$. $\square$

Proposition 4. ***Comparative Statics Without the Resource Constraint.*** *Consider the optimization problem described by equation* (23) *but ignore the resource constraint. In this model*

- $\frac{\partial N_h}{\partial c_h} < 0$
- $\frac{\partial N_l}{\partial c_h} = 0$.

Proof. These follow directly from differentiating the FOC in equation (26). $\square$

C Implications for Tax Administration

In this section, we consider the problem of sophisticated tax evasion from the perspective of the tax authority. Our goal is to understand how the tax authority responds to the adoption of concealment strategies by certain taxpayers, which we model as an increase of the cost of collecting revenue from those taxpayers by audit. We especially consider how the nature of the tax authorities resource constraints shape the response to such adoption.

C.1 Empirical Motivation

We begin with a simple empirical fact to motivate our simple model. This fact comes from taxpayers' contesting auditors' assessments, which we observe in the operational audit data. The tax assessment recommended by an auditor is their professional determination of the tax due given the taxpayers circumstances and the applicable tax laws, regulations, and revenue procedures. If the audited taxpayer (or their advisor) has a different interpretation of tax law, they can formally contest the assessment.[71] If the IRS and the taxpayer subsequently fail to reach an agreement, the case must be finally resolved in court. In complex circumstances, the resulting litigation can take several years. Public data on such disagreements can be found in IRS (2020), Table 18.

Figure A16 depicts the share of assessed tax with which taxpayers disagree with their initial assessment (before negotiation), and the share of audited taxpayers who disagreed with their assessments. We observe that the share of tax dollars assessed that is subject to disagreement hovers around 25% through the bottom 90% of the income distribution, and then increase substantially in the top 10%, up to more than 60% in the very top bin. Individuals in the bottom bin, which includes those taxpayers with negative income, disagree at comparable rates to the very top bin, reflecting that audited individuals with negative reported income are typically high-wealth individuals. The share of taxpayers who disagree follows a similar pattern, but the overall share is significantly lower. That the dollar share is much larger than the taxpayer share implies that, perhaps unsurprisingly, the very largest assessments are typically the subject of disagreements.

We interpret this evidence as jointly informative about 1) the sophistication of avoidance/evasion strategies through the distribution, and 2) the extent to which taxpayers fight their assessments and attempt to negotiate. Legal experts and practitioners are well aware that audits of sophisticated high-wealth individuals can become quite complex and contentious. However the economic importance of these types of frictions for tax compliance and administration is not well-understood.[72] The magnitude of the differences in Figure A16 suggest that this is an important question. Perhaps the most salient implication of this fact for tax administration is that the high frequency of disputes and litigation makes recovering revenue from the top of the distribution via audit more costly. In the model below, we consider the implications of this notion for the allocation of audits through the distribution.

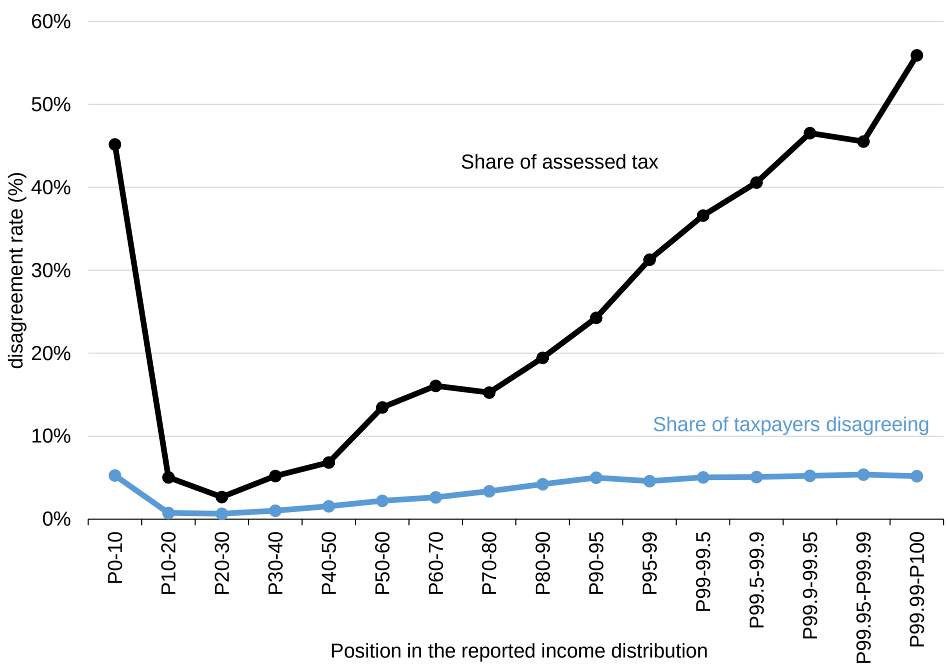
C.2 Model

Setup. There are two types of taxpayers, denoted by $\theta \in \{h, l\}$, which we basically think of as high- and low-income taxpayers. We first consider a revenue maximization problem with an exogenous resource constraint B. The tax authority decides how many of each type to audit which

[71] The same is true of random audits. In both the NRP and our treatment of operational audits above, we use the initial assessment of the auditor, before any disagreement. The two measures considered in Figure 2 are therefore comparable in this respect.

[72] See Blumenthal et al. (1998) for a model of audits as negotiations that may be relevant here.

FIGURE A16: CONTESTED AMOUNTS FOR OPERATIONAL AUDITS



Note: The top series of this figure plots the share of the total initial audit tax assessment that is contested by the taxpayer, across the income distribution. We rank taxpayers according to reported income in the tax year for which the taxpayer is under audit. The bottom series plots the share of audited taxpayers that contest their assessment amount. The data are pooled for fiscal years 2007-2018. The contested rate is very stable until the 90th percentile where it begins to increase and then rises sharply within the top 0.01% (up to 60%). The assessment share is significantly larger than the share of contesting taxpayers signifying that those with higher assessed values are more likely to contest. The large contested shares in the bottom of the distribution are mostly from taxpayers claiming large losses that are disallowed upon audit.

we denote by N_θ. Expected revenue raised by each type as a function of the number audited is $R_\theta(N_\theta)$. There is a constant marginal cost of auditing each type, c_θ. The objective is to maximize expected revenue net of costs.

The key difference between this model and the one we contrast it to later on is that we assume the total cost of audits cannot exceed some exogenous resource constraint B.

$$\max_{N_h, N_l} \quad R_h(N_h) + R_l(N_l) - c_h N_h - c_l N_l, \tag{23}$$

$$\text{subject to } c_h N_h + c_l N_l \leq B \tag{24}$$

Note that because of the presence of the resource constraint, this model is isomorphic to one in which the tax authority maximizes gross recovered revenue $R_h(N_h) + R_l(N_l)$ subject to the same resource constraint. The resource constraint requires that the last two terms in the objective function in equation (23) add up to a constant, so these terms become irrelevant for optimization.

This problem differs from an "optimal tax systems" approach to this question (Slemrod and Yitzhaki, 2002; Keen and Slemrod, 2017), in two important ways. Most importantly, the tax authority is given an exogenous resource constraint rather than simply maximizing net revenue, which we relax later. Additionally, for simplicity, we do not account for distortions induced by changes in audit policy that can cause the optimal policy to deviate from revenue maximization, such as compliance costs. Accounting for such distortions would not change the main result of interest here.

The first-order condition for an interior optimum of this problem is

$$\frac{R_h'(N_h)}{c_h} = \frac{R_l'(N_l)}{c_l}. \tag{25}$$

Comparative Statics of the Resource-Constrained Model. *In the optimization problem described by equation* (23)*,*

- $\frac{\partial N_h}{\partial c_h} < 0$
- $\frac{\partial N_l}{\partial c_h} > 0$ if and only if $-N_h R_h''/R_h' > 1$.

That increasing c_h decreases N_h is unsurprising. More interesting is that in this model, the change in c_h has an effect on audits of low-income individuals. Because the tax authority is allocating finite resources to these two types of audits, the change in c_h has two effects on N_l, which are exactly analogous to an income and substitution effect in consumer choice theory. First, holding N_h fixed, increasing c_h leaves fewer resources available for audits of type l, which tends to decrease N_l: the income effect. Second, increasing c_h induces the tax authority to substitute toward auditing more type l taxpayers.

Which one of these effects dominates depends on the curvature of the revenue function for type h, $-N_h R_h''/R_h'$, which determines whether total expenditure on h type audits goes up or down.

We next show that if we relax the exogeneity of the resource constraint, the spillover effect of an increase in c_h on audits of type l taxpayers disappears. Ignoring the resource constraints, the objective in (23) has simple first-order conditions that equate marginal revenue and marginal cost:

$$\begin{aligned} R_l' &= c_l \\ R_h' &= c_h. \end{aligned} \tag{26}$$

Comparative Statics Without the Resource Constraint. *Consider the optimization problem described by equation* (23) *but ignore the resource constraint. In this model*

- $\frac{\partial N_h}{\partial c_h} < 0$
- $\frac{\partial N_l}{\partial c_h} = 0.$

Proposition 8 states that without an exogenous resource constraint, the spillover effects from an increase in c_h onto low-income types no longer occurs in this model.[73]

Contrasting Proposition 7 and 8 helps us understand how increased concealment effort by high-income taxpayers might affect low-income taxpayers, which we view as interesting given recent debates about the allocation of resources to various types of audits. The resource-constrained version of the model is closer to how tax administration works in the real world, where the IRS is given a budget by Congress and allocates these resources toward various types of enforcement. In this model, because the tax authority is devoting limited resources to all types of audits, increased concealment effort by high-income taxpayers can actually cause the tax authority to *substitute* toward auditing more low-income taxpayers, or it can deplete resources and cause fewer audits of low-income taxpayers. The unconstrained version of the model is closer in spirit to a model of optimal policy—subject to the caveats described e.g. by Slemrod and Yitzhaki (2002). The results for this version of the model model imply that increased concealment has no impact on *socially optimal* audit policy toward low-income individuals.

[73] Key for this result to obtain is that R_h does not depend on N_l and vice versa. This seems realistic, but it could be violated, for example, if auditing one type could lead to the discovery of information that is useful for auditing the other type.