Comment on “Income Inequality in the United States: Using Tax Data to Measure Long-Term Trends” by Auten and Splinter*

John Iselin, University of Maryland Daniel Reck, University of Maryland

July 9, 2024

Abstract

We assess recent estimates of the dynamic evolution of the top 1% share of National Income in Auten and Splinter (forthcoming, *JPE*) and contrasting estimates by Piketty, Saez, and Zucman (2018, *QJE*). We focus on their largest source of disagreement: unreported income due to non-compliance. Based on our review of the methods employed by Auten and Splinter, we argue that their results are driven by the assumption that, despite the fact that 2/3 of the estimated total under-reporting is unobserved in random audit data, detection of noncompliance in random audit data is equally good in the bottom 99% as in the top 1%. We estimate that this assumption reduces the growth in the share of pre-tax national income held by the the top 1 percent by 0.46 percentage points. We review empirical evidence that total under-reporting is likely more concentrated in the top 1% than audit-detected under-reporting, with the most important source of bias coming from unreported pass-through business income, which is not systematically captured in individual random audit data. Turning to dynamics, we argue that the rise of pass-through business forms since 1986 likely causes this bias to grow significantly over time, leading to an under-statement of the increase in inequality. Other sources of potential dynamic bias include changes in random audit procedures, changes in IRS methodology for estimating un-detected under-reporting, and changes in offshore tax evasion. An apparent error in the allocation of excess depreciation exacerbates this dynamic bias.

We also review related methods by Piketty, Saez, and Zucman. These authors' methods are rooted in the assumption that if reported income grows in concentration, then so should unreported income. We propose a simple alternative method that explicitly imposes this via a distributional neutrality assumption. Under this assumption, growth in the top income share increases by 0.52 pp; Piketty et al's estimates feature an additional 0.6 pp increase due to a reranking effect. As such, conceptual disagreement about the distribution of unreported income accounts for about 60% of the disagreement between these papers; the rest is due to the reranking effect.

*We thank Gerald Auten, Judy Hellerstein, Ethan Kaplan, Melissa Kearney, and and Gabriel Zucman for their helpful feedback on this comment.

1 Introduction

A recent paper by Auten and Splinter (2023) [henceforth AS] challenges the widely held consensus that income inequality in the United States has risen dramatically since the 1980s. Influential estimates from Piketty and Saez (2003) [PS], using data from individual income tax returns, suggest the share of income received by those in the top 1% of the distribution has risen by some 9.2 percentage points over the last 60 years, from 8.4 percent in 1960 to 17.6 percent in 2019. However, these estimates are based on income reported on individual tax returns, leaving open the possibility that the true distribution of income might be different if we account for additional types of economic income. More recent estimates by Piketty, Saez and Zucman (2018) [PSZ] attempt to overcome this issue by using an expanded income concept, National Income, as measured by the National Income and Product Accounts (NIPA), and obtain substantially similar results to PS regarding the evolution of the top 1% share of income. Adopting this expanded income definition required that the authors use non-tax administrative and survey data, alongside a range of assumptions, to transition from income observed on tax returns to National Income. In their paper, AS approach the exact same research question as PSZ with similar data, but with a different set of assumptions and supplementary datasets. AS arrive at a different conclusion: that the increase in income inequality has been far more modest. They estimate that pre-tax income in the top 1% rose by 3.0 percent of National Income between 1962 and 2014.

These authors have engaged in a protracted back-and-forth (Piketty, Saez and Zucman, 2023; Splinter, 2023*a*,*b*). There are multiple sources of disagreement in methods between AS and PSZ, which makes makes tracking the debate from the outside very difficult, even for experts. In this comment, we thoroughly examine the component of National Income that generates the most quantitatively significant difference between the estimates in these studies: income that should have been included on tax returns but was not due to tax non-compliance. The differences in methods for treating unreported income drive about 1.6 of the 4.6 percentage point difference between studies in the change in the top 1% pre-tax income share since the 1960s, more than any other methodological difference (see Auten and Splinter (2023), Table 4).[1] In this comment, we focus on the disagreement around misreporting, and do not engage with the other methodological differences, nor with differences in these authors' estimates of inequality in income after taxes and public spending.

Where there is a gap between total income of some type (e.g. non-farm proprietor income) between NIPA and tax data, due to accounting for misreporting or some other reason, PSZ allocate the additional income in proportion to reported income for that type of income (and they assign no additional income where reported income is negative). AS, meanwhile, use selected moments from random audit data[2] to calibrate a micro-simulation of the distribution of unreported income, which they use to distribute the sum of the gap between NIPA and tax-return based measures of wages, rental income, farm, and S-Corporate income, as well

[1] In their published paper, AS group non-compliance-derived misreporting together with legally exempt business income (LEBI), e.g. business income that is not taxed due to accelerated depreciation allowances in many of their tables and figures. However, the 1.6 percentage point figure above reflects solely the effect of switching from AS to PSZ's treatment of misreporting (see page 31 of the online appendix of Auten and Splinter (2023)).

[2] In this paper, we refer to NRP and TCMP stratified random audits colloquially as "random audit data." This label is common and highlights a key virtue of these data: they are representative random samples of audits.

as the portion non-farm proprietor income specifically estimated to be from misreporting in NIPA. Fundamentally then, evaluating the accuracy of each assumption requires determining if unreported income is distributed like audit-detected misrreporting, or whether an allocation proportional to reported income is more accurate.

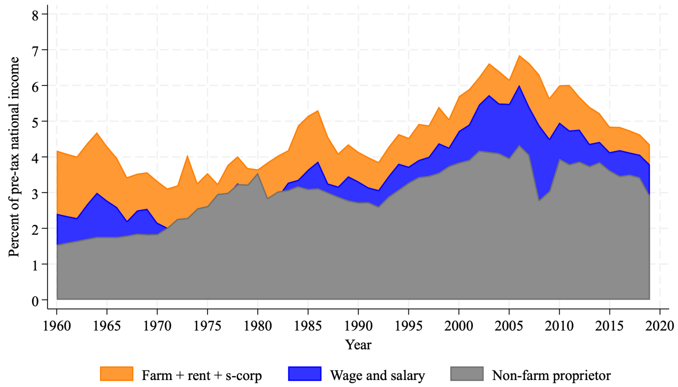
To build understanding of a key conceptual challenge here, Equation (1) decomposes total under-reported income into three components. Some under-reporting is detected in random audit data, which is the data underlying the micro-simulations in AS. National Income also includes an estimate of under-reporting that was not detected during random audits, which is identified by a method called Detection Controlled Estimation (DCE), and which is estimated to comprise about two thirds of total under-reporting in IRS Tax Gap estimates (Guyton et al., 2023*b*). The counterfactual underlying DCE is one in which all auditors are replaced by the auditors who detect the most under-reporting, so DCE does not capture another potential component of under-reporting: that which is undetectable given the information available and the audit procedures. This third component is excluded from NIPA. We include an error term to reflect that these components may be mis-measured in practice.

$$\text{Under-Reporting (UR)} = \overbrace{\underbrace{\text{Audit-Detected UR}}_{\substack{\text{Covered in}\\ \text{Random Audit Data}}} + \text{DCE-Identified UR}}^{\text{Included in NIPA}} + \text{Undetectable UR} + \text{Error} \tag{1}$$

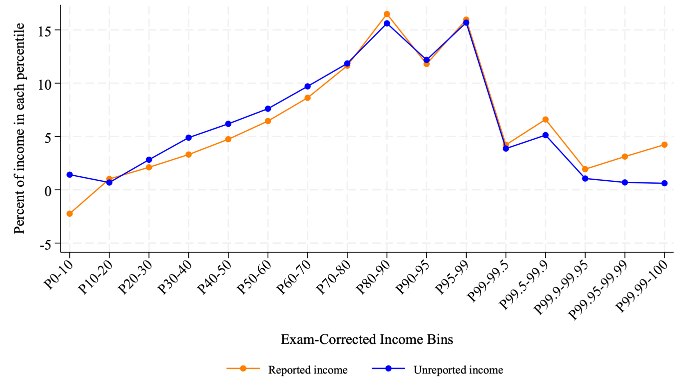
AS use the distribution of audit-detected under-reporting to allocate all under-reporting included in NIPA, and some gaps between NIPA and tax data totals that may or may not be due to under-reporting. The validity of this approach depends on some assumptions: NIPA totals for unreported income must be correct and consistently estimated over time, the DCE-identified component must be distributed similarly to the audit-detected under-reporting that is visible in random audit data, and undetectable under-reporting should not matter quantitatively. If we adopt these assumptions, then Auten and Splinter's approach has clear advantages over PSZ's. If not, we must carefully weigh what approach might be the most reasonable given the available evidence.

Figure 1: Summary: Unreported Income and the Dynamics of Inequality

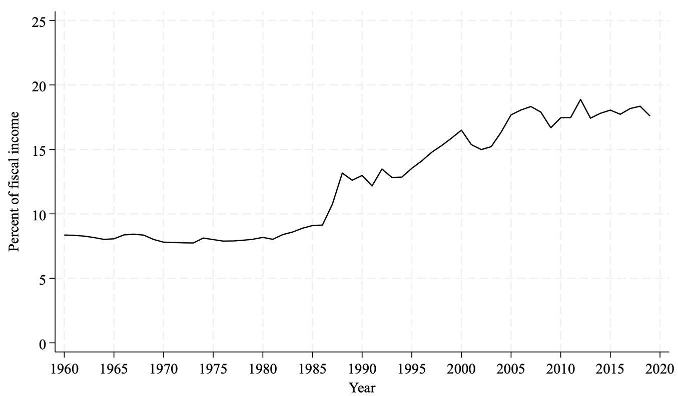
(a) Misreported income as a % of National Income



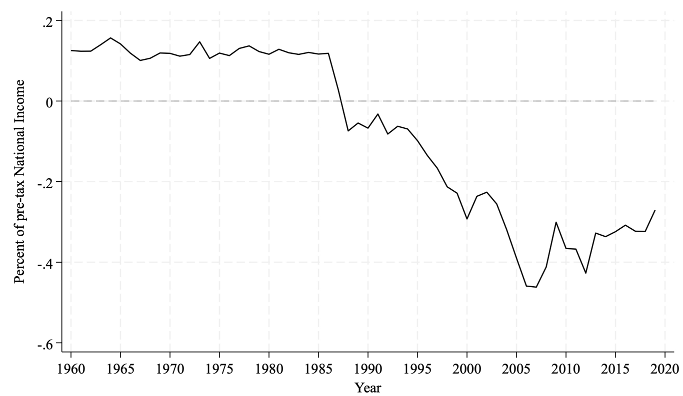
(b) Shares of reported & audit-detected under-reported income (2006-2013)



(c) Share of reported fiscal income top 1% (in %)



(d) Top 1% share of income using Auten and Splinter (2023) assumptions less reported income share (in %)



Notes: Panel (a) is derived from explicit NIPA adjustments for Wages and salaries and non-farm proprietors' income, and the implicit NIPA gaps described in Section 3 for farm, rental income, and S-Corp income. The data come from Auten and Splinter (2023), Table T-T1. We observe that the share of unreported income in National Income has grown over time and most of the growth is driven by unreported non-farm proprietor income. Panel (b) is derived from statistics from 2006-2013 National Research Program (NRP) individual random audit data reported in Guyton et al. (2023*b*). We observe that under-reporting detected in random audits is less concentrated in the top 1% than reported income. Panel (c) plots the increase over time in the top 1% share of reported fiscal income from PS, drawn from Auten and Splinter (2023), Table F-6. Panel (d) illustrates the effect of adding unreported income to national income under the assumption that national income sans misreporting and reported fiscal income exhibit similar concentration. We plot the difference between the top 1% national income share with and without including unreported income, using a macro approach based on the assumptions of Auten and Splinter (2023). The macro-approach follows Equation (2), where M_{1t} is fixed at 11.4 percent, based on calculations in Guyton et al. (2023*b*).

Figure 1 provides a high-level overview of how Auten and Splinter's assumptions shape their estimates. Figure 1a presents the totals over time of the under-reported income whose concentration in the top 1% is the source of disagreement, as a share of National Income. For "wage and salary" and "non-farm proprietor" income, we are plotting estimates of total income under-reporting included in NIPA and drawn from IRS tax gap studies. The third category is the gap between NIPA totals and IRS tax data totals for income from farm, rents, and S corporations. The NIPA totals for these three sources of income are not drawn from IRS tax data, so it is ambiguous whether this third category actually reflects under-reporting from non-compliance or some other reason that IRS tax data and other data sources disagree in the aggregate. The largest category of under-reporting is "non-farm proprietor income," which includes income from both sole proprietorships and partnerships. Total estimated under-reported non-farm proprietor income grew from about 1.5% of national income in 1960 to over 3% in 2015. To what extent this growth is driven by increases over time in unreported income for the top 1% or the bottom 99% (or biases from inconsistent measurement) will be quantitatively important for the estimated dynamics of the top income share.

Figure 1b shows that in recent years, under-reported income detected in random audits is less concentrated at the top of the income distribution than reported income. Based on data from random audits (without DCE adjustments) circa 2006–2013, 11.4% of under-reported income belongs to the top 1% by audit-corrected income, while 20.1% of reported income belongs to the top 1% by reported income. Ranking tax units by audit-corrected income, under reporting hovers around 4-5% of income in the bottom 99% of the distribution, and it falls sharply in the top 1% (see also Guyton et al., 2023*b*, Figure 1). Although it is not explicit from their description of their methods, our review of AS' methodology highlights that a core assumption of their approach is that DCE-identified under-reporting has the same concentration in the top 1% as audit-corrected under-reporting. In other words, their methods impose that the share of under-reporting belonging to the top 1% in random audit data from the blue series in Figure 1b must be very similar to the top 1 percent's share of all under-reported income in NIPA, including the DCE-identified component (see equation (1)).

Figure 1c reports the well-known estimates of Piketty and Saez (2003) (extended through 2019), showing that the top 1% share of income reported on tax returns has grown over time. The jump in this share right after 1986 is driven by income shifting between C corporation and pass-through business forms in response to the Tax Reform Act of 1986, implying that real income inequality did not suddenly jump in 1986 (Gordon and Slemrod, 2002; Alstadsæter et al., 2023). Interestingly, a relatively small portion of the dynamic divergence between PSZ and AS appears to involve how one accounts for the income of C corporations to correct this issue (see AS Table 4). They do differ significantly in how accounting for the component of National Income depicted in Figure 1a matters for the top 1% income share.

If income inequality did not actually increase very much over this period, as AS claim, there must be some countervailing component of income, which grows more rapidly in the bottom 99% than in the top 1%. AS do not directly show that unreported income in the bottom 99% has this property using, e.g., an analysis of random audit data from different periods.[3] Instead, they

[3] Doing so would be complicated not only by DCE but also by the fact that the design of the random audit program has changed substantially over time (see Brown and Mazur, 2003). This same issue also creates some concerns about the

essentially infer that unreported income must have this property due to the distinct pieces of information in Figures 1a and 1b: estimated total unreported income is a growing component of National Income over time, and the detected component of under-reported income is estimated to mainly belong to the bottom 99% in random audit data.

In Figure 1d, we implement a simple procedure that mimics the methods of AS. To illustrate the effect of the allocation of unreported income in isolation, we suppose that apart from the misreporting component, pre-tax national income is distributed similarly to reported fiscal income. Denoting the top 1% reported income share (from Figure 1c) by R_{1t}, and the share of national income misreported (from Figure 1a) as M_t, we assume a fraction $R_{1t} * (1 - M_t)$ of national income belongs to the top 1%. For the under-reporting component of national income, we assign to the top 1% a fraction $M_t * E_1$ of national income, where $E_1 = 11.4\%$ is the share of audit-detected misreporting belonging to the top 1% (by audit-corrected income) from the blue series in Figure 1b.[4] The remainder, $(1 - E_1)M_t$ is allocated to the bottom 99%. Altogether the estimated share of national income belonging to the top 1% Y_{1t} is therefore:

$$Y_{1t} = R_{1t} * (1 - M_t) + E_1 * M_t. \tag{2}$$

This exercise mimics the methods of AS in a simpler way (without accounting for other types of income absent from tax returns like C corporation income). We justify this method, alongside some alternatives and validation exercises, in Section 5.2. In Figure 1d, we plot $Y_{1t} - R_{1t}$, which we can interpret as the influence of AS' assumptions compared to an estimate based on the concentration of reported income, like PS. The results suggest that AS's allocation of unreported income decreases the estimated level of the top income share by about 0.3-0.4 percentage points in the last decade of the series, and it decreases the estimated change in the top 1% income share over time by about 0.46 percentage points over the full period.

We argue that the assumption that all under-reporting is distributed like audit-detected under-reporting is unrealistic. AS do not offer any evidence in support of this assumption for components of under-reporting not covered by random audit data.[5] Furthermore, we argue that there are multiple specific threats to the validity of this assumption. Empirical data suggests that using individual random audit data to allocate under-reporting is likely to lead to bias in an estimate of the top 1% income share over time. First, and most importantly, pass-through income is about a third of all income in the top 1% of the income distribution, and Guyton et al. (2023*b*) documents that due to audit procedures (and potentially resource constraints),

comparability of the total figure used in NIPA over time, which we discuss below. Auten and Langetieg (2023) analyzes data from different waves of random audit study and creates tables that are used to calibrate the micro-simulations of AS, but because the focus is on accounting for re-ranking rather than analyzing the dynamics of under-reporting directly, we find it impossible to draw any conclusions about the dynamics of under-reporting and inequality from these results. We discuss this further below.

[4] In this illustrative exercise, we assume the share E_1 is fixed over time. The micro-simulations used by AS draw on different waves of random audit data, as we discuss below. From the tables in Auten and Langetieg (2023) upon which the AS micro-simulations are based, we do not know how the share E_1 differs across waves. The extent to which any changes in E_1 over time are driven by inconsistent measurement (changes to audit procedures, the rise of pass-throughs) is not discussed in Auten and Langetieg (2023).

[5] We discuss in more detail in Section 4.1, but the evidence cited by AS in Appendix Figure B5 speaks the viability of their micro-simulation approach as a means of approximating the distribution of audit-detected misreporting, not that detected and DCE-identified (or undetectable) misreporting share the same distribution.

this income is not examined in the random audits that AS treat as representative of all types of under-reporting including pass-through under-reporting. Second, regarding DCE-identified under-reporting (about 2/3 of the NIPA total), complexity of income is markedly higher at the top of the income distribution. If the depth of examination by an auditor is correlated with what they detect when examining complex returns, the auditor effects captured by DCE should be more concentrated at the top than detected under-reporting. Third, offshore tax evasion is almost completely undetected in random audits and concentrated at the very top of the income distribution (Guyton et al., 2023*b*, show this for the US and several studies document similar patterns in other countries). Available evidence suggests that this was substantial, at least prior to a recent wave of enforcement. These first three concerns are overlapping in the sense that some of what is identified by DCE might be top-end, sophisticated under-reporting in offshore or pass-through structures that would not be detected by less thorough auditors.[6]

Turning to dynamics, the most obvious threat to AS's estimates comes from the dramatic increase in the importance of pass-through business structures, especially partnerships, during this period (Smith, Zidar and Zwick, 2022). This change, commonly called the "rise of pass-throughs," is arguably the most important change to the taxation of business income that occurs during the period of interest. The aforementioned bias due to AS' assumptions about pass-through under-reporting is likely to grow substantially over time as pass-throughs grow in importance. While it is likely less quantitatively important than the rise of pass-throughs, the rise of offshore evasion during the decades preceding a crackdown in 2008 could also lead AS to under-state the increase in inequality over this period, while any reductions in offshore tax evasion more recently could have the opposite effect since 2008. There here have also been changes over time to random audit procedures and IRS Tax Gap estimation methods that could cause inconsistencies over time in NIPA's total figures for unreported income. We review these changes below; their quantitative importance is unclear.

In contrast, PSZ allocate unreported income at the micro level in proportion to positive reported income for each type of income. This approach imposes that if reported incomes of some type become much more unequally distributed – which occurs during the period of interest due to the rise of pass-throughs – then unreported income of that type will have the same property. AS point out that PSZ's micro allocation could bias upwards their estimates of the concentration of income in the top 1% due to re-ranking effects, especially given that they assign no under-reporting to those with reported losses. We propose a simple alternative method in which we allocate under-reporting similarly to Equation (2), but we break total under-reporting (M_t) into components by type of income. Using public Distributed National Accounts (DINA) micro-data produced as part of Piketty, Saez and Zucman (2018), we distribute these according to the

[6]This overlap is the subject of much discussion and sensitivity analysis in Guyton et al. (2023*b*), because it creates a methodological challenge about accounting separately for undetected sophisticated under-reporting and the under-reporting identified by DCE. In a comment on an earlier version of Guyton et al. (2023*b*), AS take issue with the the authors used to address this overlap (Auten and Splinter, 2021). This methodological debate is irrelevant for the point at hand. Our argument here is that either DCE adjustments do not account for sophisticated evasion at the top, in which case we have unaccounted-for top-end under-reporting, or DCE adjustments do account for sophisticated evasion at the top, in which case too little of the DCE adjustments is allocated to the top by AS, because they impose that DCE-identified under-reporting is distributed like detected under-reporting. Either way, we have a downward bias in the top 1% income share.

share of income of each type reported by the top 1% in reported income data.[7] Our benchmark imposes distributional neutrality by type of income, which capture the idea that growth in inequality of reported income suggests unreported income is also increasingly unequal. Moreover, building on Guyton et al. (2023*b*), we make this assumption directly at the aggregate level rather than attempting to capture it via a micro-level allocation, which addresses AS' criticism about excess re-ranking from the PSZ method.

Because the growth in unreported income in Figure 1a is driven by relatively concentrated types of income (mainly non-farm proprietor income), and the concentration of these types of income has actually grown over time (mainly due to the increase in the importance of pass-through income at the top), our benchmark series suggests that accounting for unreported income increases increases the top 1% share of income by about 0.52 percentage points from 1962 to 2014. This increase is 0.63 percentage points smaller than the increase implied by the approach of PSZ, suggesting that their allocation at the micro level may generate some bias from excess re-ranking (e.g. by not allocating under-reporting to those with business losses). We emphasize that whether there is really a bias here and its quantitative magnitude depends on empirical questions that we cannot answer confidently with the available data (e.g. without random audit data that comprehensively captures misreporting in pass-through businesses).

Finally, we make a subtler comment about accounting for excess depreciation. We observe that Figure 1d differs from Appendix Figure B6 of AS; this difference arises primarily because the AS Appendix figure combines adjustments for under-reporting with a separate adjustment for excess depreciation. The former decreases the top income share; the latter adjustment offsets this decrease, but it also appears to be incorrectly specified by AS.

By way of background, capital depreciation contributes negatively to economic income by definition, and the US tax code allows businesses to deduct their capital expenses over time according to a depreciation schedule. However, the code allows businesses to deduct capital expenses more rapidly than economic depreciation, and the generosity of this excess depreciation allowance grew over the period of interest to us, as this became a popular tool to attempt to stimulate investment (see e.g. Zwick and Mahon (2017)). To correct for excess depreciation and estimate economic rather than taxable income, NIPA includes an extra component of business income it calls the capital consumption component. An early draft of AS incorrectly treated this component of income as a part of under-reporting, so that it was allocated primarily to the bottom 99% of the distribution based on random-audit data. The published version of AS does not make this mistake, but instead they allocate the excess depreciation component of "non-farm proprietor income" *in proportion to the depreciation of sole proprietors, not sole proprietors and partnerships*.[8] This is a consequential error, because partnership income is much more con-

[7]We use the February 2022 vintage of the DINA files, downloaded from https://gabriel-zucman.eu/usdina/. As noted in their codebook, these files are not identical to those used by PSZ to calculate their 2018 results. Rather, each observation in the DINA files is a synthetic individual, but they do allow us to approximate the results of both PSZ and AS, and expand on their methodologies.

[8]The relevant line of their code is:

```
UNDER = UNDER + 0.85*PropExpn + (1000000*&&nfcc&yr - 0.85*totexp)*(PropDepr/tPropDepr) +
1000000*&&nfres&yr*(pPARTSCP + pBUSN)/(ppartscptot+ptotbusn);
```

This indicates the capital consumption/excess depreciation adjustment for both sole proprietorship and partnerships ("nfcc") is allocated proportionally to depreciation deductions of sole proprietorships only ("PropDepr"). This issue was pointed out to us by Gabriel Zucman when we approached him with some questions about the methods in PSZ for centrated in the top 1% of the income distribution than sole proprietorships, and partnerships claim substantial excess depreciation.

Quantitatively, total excess depreciation for non-farm proprietors was $206 billion in 2014, or 1.4 percent of National Income. About 16% of sole proprietor income belongs to the top 1% compared to about 46% of sole proprietor plus partnership income. If about 46-16=30% of excess depreciation income is misallocated to the bottom 99%, the top 1% share would be too low by about 30%*1.4% = 0.41 percent of national income – enough to fully offset the decrease we found in Figure 1d and therefore to reconcile our findings with AS Appendix Figure B6.[9]

In summary, we find that differences in the assumptions made about the distribution of misreported income explain 1.0 percentage points out of the 1.6 percentage point gap between AS and PSZ's estimates of the rise in the top 1 percent's share of pre-tax national income between 1962 and 2014. Combined with potential excess re-ranking due to the micro allocation in PSZ (about 0.6 pp), we can fully explain the quantitative divergence between AS and PSZ attributable to unreported income. The remainder of this comment provides a more thorough and technical version of the argument we summarize above. Section 2 reviews the conceptual challenges in estimating under-reporting at the distributional level. Section 3 reviews AS's methods for handling under-reporting in detail. Section 4 reviews the methodological issues and empirical evidence suggesting that AS's approach leads to significant bias in the estimated dynamics of the top 1% income share. Section 5 turns to alternatives to AS's approach, including that of PSZ and our proposed alternative. The final section concludes.

2 Conceptual Challenges

We begin by we reviewing the underlying methodological and conceptual challenges that are inherent to this exercise.

At face value, using random audit data to calibrate a micro-simulation of under-reporting has a lot of intuitive appeal. The perfect dataset to perform this exercise would be a random sample of tax returns in which we could observe reported and unreported income perfectly. While the data from IRS random audit studies are indeed (stratified) random samples, there are four challenges to the use of these data for the exercise at hand.[10]

Challenge 1: Re-ranking. The first challenge is accounting for re-ranking effects. If we wish to estimate the share of income belonging to the top 1% of the *true income distribution*, we must account for the fact that the set of tax units in the top 1% of the true income distribution is different from the set of tax units in the top 1% of the reported income distribution. In fact, the top 1% of the reported income distribution are negatively selected on non-compliance, and random audit data suggest that most of the non-compliance belonging to the top 1% by corrected income comes from tax units that are not in the top 1% by reported income, but who move into

this comment. We confirmed the purpose of this line of code, and that partnership depreciation is not included in the underlying data. The latter is unsurprising, because the SOI data AS use is a sample of Forms 1040 that should contain information about the deductions of sole proprietors (Form 1040 Schedule C) but not partnerships (Form 1065).

[9] In 1962, NIPA's capital consumption adjustment was $-100 million, less than 1 percent of the 2014 dollar amount: therefore, the change in the top 1 percent share is captured by the 0.41 figure above.

[10] More specifically, NRP random audit data are a stratified random sample that over-samples self-employed and high-income taxpayers to increase precision, with corresponding sampling weights.

the top 1% after re-ranking by corrected income.[11]

If re-ranking were the only challenge here, it would not be difficult to handle. One would just need estimates of the joint distribution of reported income and unreported income, so that the latter could be calibrated accurately in micro-simulations using a dataset in which only reported income is observed. This is the main goal of Auten and Langetieg (2023), discussed further below. However, accounting for re-ranking becomes much more difficult when we confront the next two challenges.

Challenge 2: DCE-Identified Under-reporting. It has long been recognized that random audit data do not capture all under-reported income. This is why the preferred estimates from the IRS employ Detection Controlled Estimation. According to current DCE methods, the random audit data circa tax years 2008–2013 fail to capture 66% of total estimated under-reported income.[12] The DCE methods are rooted in models of auditor effects. The conceptual idea, similarly to what we find in the literature on Teacher Value-Added (Chetty, Friedman and Rockoff, 2014*a*,*b*), is to estimate detected under-reporting in a counterfactual where all auditors are replaced by auditors estimated to be the most effective at detecting under-reporting.

The methodology of DCE leaves room for doubt about the total estimate of undetected under-reporting, its dynamic evolution, and its distributional properties (e.g. how much undetected under-reporting belongs to the top 1%). In order to separately identify auditor effectiveness from underlying non-compliance, these methods require the assumption that auditor assignment to specific returns is "as good as random," but in practice auditors are assigned based on their expertise and the sophistication of the return (Guyton et al., 2023*a*, confirms this from a review of random audit case allocation procedures). Where there is uncertainty about how tax law applies to a complicated tax position, DCE effectively defers to the most aggressive auditors' interpretation of tax law (Hemel, Holtzblatt and Rosenthal, 2021). The specifics of DCE methods used in IRS Tax Gap studies have also changed over time ((Feinstein, 1990, 1991; Erard and Feinstein, 2010, 2011), see Guyton et al. (2023*a*) for a discussion of the differences in methods). This raises questions of comparability over time for both total and distributional estimates (note that the former feeds directly into NIPA).

These DCE methods were also designed primarily to estimate population-level totals, rather than for distributional analysis. The micro-simulations of the underlying DCE model the IRS uses to map total income under-reporting to the total tax gap in the Tax Gap studies imply a distribution of undetected under-reporting. However, these methods have changed significantly over time,[13] and virtually all researchers who have engaged with these distributional estimates acknowledge that the distributional features of the methods are imperfect (Johns and Slemrod, 2010; Jason Debacker et al., 2020; Auten and Langetieg, 2023; Guyton et al., 2023*a*). The main difficulty with distributional DCE comes from re-ranking effects: properly accounting

[11]For empirical illustrations, see Figure 3d below, as well as Guyton et al. (2023*b*) Figure A3 and Table A2 and the similar figure in Auten and Langetieg (2023).

[12]This figure is calculated based on using estimates from Guyton et al. (2023*b*) Tables A1 and A6; it is exact up to rounding error and statistical uncertainty. See also Johnson and Rose (2019). The precise figure varies over time and is not available in most IRS publications of tax gap estimates – only estimates including DCE are typically reported.

[13]The methods used for the Tax Gap studies on NRP random audit from tax year 2001 (Black et al., 2012) are described in Feinstein (1990). The methods used in the tax gap studies for tax year 2008 and beyond (Johnson and Rose, 2019) are described in Erard and Feinstein (2010). See Guyton et al. (2023*a*) for additional discussion of how these methods have changed and implications for distributional analysis.

for re-ranking by true income, as opposed to reported or audit-corrected income, requires estimates of the joint distribution of reported income, audit-detected under-reported income, and undetected under-reported income. Identifying this joint distribution – or identifying enough about it to estimate the top 1% income share by true income – requires additional assumptions beyond those necessary to identify total undetected under-reporting. Micro-simulations that include under-reporting that is identified by DCE methods must confront this problem.

Challenge 3: Undetectable Under-reporting. Setting the challenges about how to incorporate DCE aside, it is also unclear whether DCE does an equally good job of accounting for undetected under-reporting throughout the income distribution. Conceptually, if the method is valid, we can think of DCE as estimating under-reporting that could have been detected given the information available to auditors and audit procedures, but which is not detected in practice due to imperfections in auditors' use of information and execution of audit procedures. In other words, DCE identifies *detectable but undetected* under-reporting. There may also be *undetectable under-reporting* that even the best auditors cannot detect given their information and the procedures they must follow. Moreover, because they are audited more often, undetectable forms of evasion would be especially attractive to high-income taxpayers.

Guyton et al. (2023*b*) suggest that at least two forms of under-reporting are undetectable in the most recent random audit data and quantitatively significant at the top of the income distribution: under-reporting of offshore financial income, and under-reporting of pass-through income. A wide body of research documents the importance, in the US and elsewhere, of tax evasion that makes use of offshore intermediaries (e.g. concealed foreign financial accounts). Guyton et al. (2023*b*) find that this type of evasion is virtually never detected in random audits (which is perhaps unsurprising given the international secrecy involved). Data generated by a recent crackdown on offshore evasion, however, suggest that offshore evasion was prevalent at the very top of the income distribution during the years Guyton et al. study.

Unlike offshore evasion, the lack of coverage of under-reporting in pass-through businesses in random audit data is attributable to audit procedures rather than auditors' information. When an individual owner of a pass-through business was audited during an NRP random audit, the auditor would verify that the income allocated to the business owner by the pass-through was duly reported by the individual, and that the individual complied with certain rules (e.g. excess business loss limitation rules), but they would virtually never verify that the pass-through business itself reported income truthfully.[14] Detection of entity-level under-reporting in NRP random audit data is extremely rare and under-reporting of pass-through income is consequently under-estimated. Recall pass-through income includes income from S corporations and partnerships. The latter should be included by definition in under-reported non-farm proprietor income discussed above, while S corporation income belongs in a different category.

Challenge 4: Dynamics Given that a central objective here is to estimate changes in inequality over time, and that we must make use of imperfect data in order to overcome the

[14]Our description of NRP random audit procedures is based on data from Guyton et al. (2023*b*). Description of audit procedures used in the TCMP are difficult to come by, but we believe that the same basic fact is true of TCMP random audits. For instance Slemrod (2007) states "TCMP audits of personal tax returns do not generally investigate corporate or partnership tax returns, so any evasion at those business levels is generally not accounted for."

first three challenges above, we should be particularly mindful of changes over time that might introduce bias into our estimates from any given method. Regarding challenge 2, we should exercise caution around changes in audit procedures from Taxpayer Compliance Measurement Project (TCMP) in 1988 to the NRP in 2001 (Brown and Mazur, 2003), and the aforementioned changes in DCE methods. Regarding challenge 3, we note that accounting for potentially undetected under-reporting in partnerships and the role this plays in the distribution of under-reported proprietor income is particularly important in light of the so-called "rise of pass-throughs" over time (Smith et al., 2019). Meanwhile, for offshore evasion, we should be mindful not only of the potential decline in offshore evasion after an ambitious crackdown that started in 2008, which AS discuss in some defense of their methods, but also increases in offshore evasion during the three decades of increasing global financial sophistication that preceded the crackdown (Zucman, 2013).

3 Auten and Splinter's Methodology

A core goal of AS, as well as PS and PSZ, is to determine how the distribution of income has changed over the past 60 years. Both groups of authors use the same basic approach to study this topic: begin with a dataset of tax data, employ a process of adjustments, additions, and subtractions to produce a measure of income for each household that is consistent with the income concepts reported in NIPA. We are interested in the distributional impacts of one of those expansions: the inclusion of misreported income, which is income that should appear on a tax return but is not present. This can be due either to intentional evasion or to misunderstanding, ambiguities in tax rules, or other less nefarious causes. In determining how to account for this gap in the calculation of the distribution of income, there are two critical questions: how big is the gap, and how should the missing dollars be distributed among households? A full description of AS's methodology regarding misreporting is available in Appendix A.

AS calculate the total amount of misreporting by income source and year by taking the gap between NIPA and tax-reported total income for each of five different types: wages and salaries, rental income, S-corporation income, farm proprietor income, and non-farm proprietor income. However, for non-farm proprietor income, the overall gap is larger than the explicit adjustment for unreported income due to the additional adjustment NIPA makes for excess depreciation/capital consumption, which is distributed separately (this is the source of the error discussed in the introduction). After removing excess depreciation and five percent of the remaining gap to allocate to non-filers, AS combine the totals for the five types of income, arriving at a total amount of misreporting that they distribute at the micro level to tax filers.

The procedure used by AS to determine the distribution of misreporting reflects the fact that the misreporting gap represents a mix of audit-detected misreporting and DCE-identified misreporting. Their first step is to estimate size and distribution of detected misreporting using the methodology proposed by (Auten and Langetieg, 2023, henceforth AL). From this paper, AS can obtain the ratio of corrected income (reported plus audit-dected misreporting) to reported income by year, rank of reported income, and - importantly - ratio groups corresponding to ranges of the ratio of reported and audit-corrected income. AL tabulate the share of tax returns in a given income group (and wave of random audit data) whose ratio of corrected income to reported income falls in a particular range (ex. 2-4 times), as well as the mean and standard deviation of the ratio of corrected to reported income within each ratio group.[15] AS use these statistics as part of a micro-simulation approach to distribute audit-detected misreporting to individual tax returns. Tax units are assigned to the correct income bin, then randomly assigned a ratio group, from which their corrected income – and therefore audit-detected misreporting – is derived.

After simulating values for detected misreporting, AS follow the process outlined in Auten and Splinter (2021) to simulate the size and distribution of DCE-identified undetected misreporting. Their parametric specification, referred to as distributionally consistent gradient multipliers (DCGM), starts with the amount of audit-detected misreporting calculated above. AS multiply this amount by a multiplier that varies with the tax returns ratio class from above; an assumption based on the notion that returns with large detected under-reporting would be more likely to have been assigned more effective auditors, and therefore should have smaller DCE-identified amounts of reporting. Using these ratio-specific multipliers, AS produce simulated values of each tax unit's undetected misreporting following the expression above. This is summed over all tax units to obtain the overall undetected share of misreporting.

4 Sources of Bias in Auten and Splinter's Methodology

We now return to the challenges described in Section 2 and discuss how the methods employed by Auten and Splinter may fail in addressing each of these challenges.

At a high level, we can summarize what AS do in terms of two straightforward assumptions. First, AS estimate the distribution of detected under-reporting and specify a micro-simulation based on this distribution, the parameters of which impose the restriction that *all under-reporting is distributed like detected under-reporting*. In other words, when they triple the amount total under-reporting by incorporating DCE, they assume the concentration of under-reporting does not change. This restriction on the parameters governing distributional DCE is an assumption not disciplined by any formal model or empirical data beyond audit-detected under-reporting. A second, more obvious assumption AS make is that their totals for unreported income are correct.[16] For wage and non-farm proprietor income this amounts to trusting the Tax Gap estimates on which NIPA draws for these figures; for other forms of income this amounts to attributing the discrepancy between tax data and NIPA totals (which do not draw directly on tax data or Tax Gap estimates) to under-reporting.

If both of these assumptions are correct, the micro-simulations developed by AS would not only provide an accurate estimation of the distribution of unreported income, but they would be a valuable tool for simulating under-reporting with data on reported income alone. AS provide some defense of the second assumption in the back-and-forth with critics of their paper. We do not agree with the attribution of all disagreement between NIPA totals not drawing on tax data and tax data itself to under-reporting, which decreases top income shares relative to a distributionally neutral rescaling to account for these disagreements. However, such disagree-

[15]AL report the conditional mean and a quantity labelled the "standard error for mean." AS' microsimulations seem to require conditional standard deviations; whether they calibrate these standard deviations based on unreported tabulations from AL or there is an error here conflating standard errors and standard deviations is unclear.

[16]PSZ make a similar assumption without attributing all discrepancies between NIPA and tax data to under-reporting: they assume that NIPA-reported totals are accurate, and that that income that appears in NIPA exhibits similar concentration to reported incomes, regardless of whether differences between NIPA and tax data are due to under-reporting.

ments are a small and stable share of national income over time, so the scope for bias from this concern appears to be relatively small (see the orange area in Figure 1a). In our view, the implicit assumption that all under-reporting is distributed like detected under-reporting is more important quantitatively. In Sections 4.2 and 4.3, we discuss specific issues that cast doubt on this assumption, building on our earlier discussion of the conceptual challenges inherent in the exercise. We observe that the issues discussed in these two sections are overlapping to some degree because e.g. the extent to which DCE captures undetected pass-through under-reporting is ambiguous. If most pass-through under-reporting is implicitly captured by DCE, the concerns raised in Section 4.2 should carry more weight, while if not, the concerns raised in Section 4.3 are more important.

4.1 Re-ranking

A key claim by Auten and Splinter is that the method outlined in Section 3 accounts for re-ranking effects. This is the focus of their discussion of their method, rather than the implicit assumptions highlighted above. Interestingly, both AS and PSZ in their back-and-forth have accused one another of being confused about re-ranking. In reviewing the AS specification above, we find the structure that is imposed on re-ranking opaque, especially for the DCE-identified component of under-reporting. The “distributionally consistent gradient multipliers” used to allocation this portion of under-reporting are not based on any explicit model. Neither are they “distributionally consistent” in the sense of converging to the true distribution as the sample size grows large. Rather, the label “distributionally consistent” appears to refer to the fact that the specification of these multipliers imposes that the distribution of undetected under-reporting is the same as the distribution of detected under-reporting after re-ranking. This is consistent with how AS justify their approach in both Auten and Splinter (2021) and Auten and Splinter (2023). They show that the concentration of all under-reporting in their simulation matches the concentration of audit-detected under-reporting by exam-corrected income, i.e. the concentration of under-reporting *before DCE*. We depicted this concentration in the blue series in Figure 1b. More specifically, they compare their micro-simulation results to estimates from Johns and Slemrod (2010) and Jason Debacker et al. (2020), and their own analysis of random audit data, all excluding DCE. This exercise only validates that their micro-simulation accurately captures reranking effects under the assumption that all under-reporting is distributed like detected under-reporting.[17]

4.2 Detection Controlled Estimation

Is the concentration of undetected under-reporting identified by DCE similar to that of detected under-reporting? With the information available to researchers outside the IRS, which includes random audit data but does not included auditor information or estimated auditor effects from

[17] This is stated explicity by AS: "This method produces results similar to NRP-based estimates of the distribution of underreporting in Johns and Slemrod (2010) and DeBacker et al. (2020), as seen in online appendix Figure B5." (Auten and Splinter, 2023, p. 20). The contents of Figure B5 disregard the estimates of Johns and Slemrod (2010) and Jason Debacker et al. (2020) which do incorporate DCE. Their justification for ignoring the estimates that include DCE appears to stem from their claim that the DCE multipliers used in these papers bias top 1% under-reporting upwards Auten and Splinter (2021), but as discussed in Reck, Risch and Zucman (2021), the direction of the bias with this approach is actually ambiguous.

the DCE model, we find no credible and complete answer this question.[18] One approach might therefore be to remain agnostic about the distribution of DCE-identified under-reporting and present sensitivity analysis. This is the approach taken in Guyton et al. (2023*a*) and Guyton et al. (2023*b*).[19] In the specifications entertained by Guyton et al. (2023*a*), the share of total unreported income belonging to the top 1% of the true income distribution varies from around 15% to over 30% depending on the specification, which would imply the top 1% fiscal income share falls between 18.1 and 20.5% depending on what assumptions one makes about DCE. Distributional statistics not specific to the top 1% (e.g. estimates pertaining to the top 20% rather than the top 1%) are much less sensitive: the top 1% is a relatively small group in the very top tail of the distribution and mechanically, the most egregious instances of under-reported income wind up at the top after re-ranking. Consequently, small changes in model parameters can generate a lot of statistical variation. Meanwhile, AS do not present any sensitivity analysis around DCE. What a sensitivity analysis would suggest regarding the dynamics of estimated inequality is even more uncertain.

If we take a step back from random audit data, however, we find a variety of other data sources and lines of reasoning that suggest that undetected under-reporting should be more concentrated at the top of the distribution than under-reporting detected during audits. Conceptually, the underlying question here is about the covariance between the probability that under-reporting is detected and (true) income. A number of empirical facts suggest that this covariance is negative, which would imply that undetected under-reporting is more concentrated at the top than detected under-reporting, and therefore that Auten and Splinter's estimates of the top 1% income share are biased downward.

For example, one might reasonably expect detection probabilities to depend negatively the complexity of the tax return. The tax returns of the top 1% by income are more complicated than the bottom 99% in almost every sense one can imagine.[20] A potential countervailing force involving complexity is that more individuals are involved in the preparation of a complex high-income return; this could make outright fraud more difficult (e.g. due to whistleblower risk) but it also enables the aggressive mining of grey areas in the law (Hemel, Holtzblatt and Rosenthal, 2021), some of which entails non-compliance. For better or worse, the NIPA totals for proprietor income and underlying Tax Gap estimates respect auditors' initial recommendations, so if any of the recommendations involving grey areas are more likely to be made by a high-fixed-effect auditor than a low-fixed-effect auditor, we have undetected under-reporting in NIPA that is more concentrated at the top than detected under-reporting.

Relatedly, the information available to auditors in examining the tax return has a differ-

[18]While the underlying models of DCE used in recent tax gap studies are outlined in Feinstein (1991) and Erard and Feinstein (2010), our understanding is that the estimates of auditor effects from the DCE model, or replication code and data on auditor identity that would allow other researchers to replicate the estimated models, have not been made available to researchers outside the IRS. Consistent with this, AS do not employ an estimated model including the fixed effect of the auditor assigned to each case in random audit data. Even if estimated fixed effects were available, how to adapt the underlying model for distributional analysis of top income shares is debatable, as discussed further in Guyton et al. (2023*a*).

[19]The agnostic treatment of DCE appearing in these papers was in part a response to a comment by AS, Auten and Splinter (2021), on the distributional DCE methods in an earlier version of Guyton et al. (2023*b*) which in turn were based on Johns and Slemrod (2010).

[20]We think this is common knowledge, but for an empirical illustration, refer to the complexity of business ownership structures documented in Guyton et al. (2023*b*) Figure A8.

ent character at the top of the income distribution (Guyton et al., 2023*b*, Figure 1b). Around the top percentile of the income distribution, the composition of income shifts markedly, from wage and sole proprietor income to partnership, S corp, and financial capital income. Wage income constitutes about 80% of income in the bottom 99% and this is subject to comprehensive information reporting for the vast majority of workers (the main exception being tipped workers in a narrow class of occupations). Sole proprietor income is subject to some information reporting of revenues (see e.g. Forms 1099-MISC, 1099-K, and 1099-NEC); costs are not third-party reported but the taxpayer must provide receipts upon audit for all deductions they claim. Undetected under-reporting in the bottom 99% therefore likely involves revenues or tips the auditor does not discover (e.g. from informal cash transactions) and questions about eligibility to claim various deductions, e.g. questions about business vs personal use for self-employed individuals. The forms of potentially undetected under-reporting at the top of the distribution are myriad and more difficult to enumerate. Similar margins for under-reporting available to sole proprietors are also available to pass-through businesses, but the complexity of these businesses facilitates additional possibilities for sophisticated tax evasion (see e.g. micro-captive insurance transactions). Because partnerships can be owned by other entities rather than individuals (corporations, tax-exempt entities, trusts, foreign entities, and other partnerships) complexity of partnerships in particular creates widely discussed scope for sophisticated non-compliance (Black et al., 2023; Love, 2021; Cooper et al., 2016). As discussed in the next section, virtually all entity-level pass-through under-reporting is also undetected in recent waves of the random audit data by design. Turning to financial capital income (interest, dividends and capital gains), we note that third-party reporting makes non-compliance virtually automatically detectable when the income is domestic source and received directly by the individual (e.g. due to Forms 1099-Div, 1099-Int, and 1099-B),[21] Undetected under-reporting of financial capital income is therefore likely to include income that is received via foreign intermediaries and/or pass-through businesses. Data in Guyton et al. (2023*b*) and Johannesen et al. (2023) show that such foreign-source financial capital income is highly concentrated in the top 1% of the income distribution. In the top 0.1% of the income distribution, for instance, about one quarter of fiscal income is financial capital income received via a pass-through business entity, and more than 40% of taxpayers hold at least some assets via offshore financial intermediaries; both of these figures are negligibly small in the bottom 99% of the income distribution.

Third, overall audit rates are higher at the top of the distribution (Werfel et al., 2024, Table 17). A higher audit rate deters the adoption of forms of under-reporting that are likely to be detected by auditors and incentivizes more sophisticated forms of under-reporting that are less likely to be detected (see Guyton et al. (2023*b*) for a formal argument). Therefore, we would expect the relatively high probability of being audited to shift the composition of under-reporting from detected to undetected under-reporting, holding all else fixed. This gives yet another reason to suspect that undetected under-reporting is likely more concentrated at the top of the income distribution than detected under-reporting, contrary to the assumption imposed by AS.

Finally, we note that the fact that auditors are not randomly assigned raises doubts about

[21] The Form 1099-B is not as comprehensive a source of third-party information about capital gains transactions as the Forms 1099-Div and 1099-Int are for dividends and interest, respectively. Form 1099-B reporting only covers capital gains realized via a broker, and the form did not report cost basis until tax year 2011.

the overall magnitude of estimated under-reporting including DCE-identified under-reporting. Here, we have some sympathy for the desire of AS to take official statistics at face value. We appreciate their rationale for trusting NIPA aggregates in the baseline scenario, but a better approach in our view would have been to be transparent about how much this matters. It should be straightforward, for instance, to present readers with estimates of the top 1% income share that exclude DCE-identified under-reporting from total income, which would help them understand how this component of National Income matters for the dynamic evolution of inequality in their benchmark scenario. With the information AS have shared publicly about their methods and calculations, we are unable to construct such a scenario with confidence.

4.3 Undetectable

The questions around DCE revolve around the distribution of income that is estimated to be a part of National Income. Undetectable under-reporting is a part of National Income according to the conceptual definition of economic income, but it is not a part of *estimated* National Income by construction. How might missing this component of National Income bias distributional estimates? Answering this question precisely is difficult because we have limited data on undetectable under-reporting, but the weight of the evidence suggests that missing undetectable under-reporting biases estimated top 1% income shares downward.

Estimates from Guyton et al. (2023*a*) suggest that perhaps 75% of concealed offshore wealth belonged to the top 1% of the income distribution prior to the 2008 crackdown; the ownership share is similar in more comprehensive FATCA data from more recent years, but it is less clear the extent to which offshore wealth is associated with noncompliance in recent years (Johannesen et al., 2023). Given macro estimates of the overall amount of concealed offshore wealth (about 7% of GDP according to Alstadsæter, Johannesen and Zucman (2018) and a reasonable rate of return on this wealth (about 6% according to Johannesen et al. (2023)), Guyton et al. (2023*b*) illustrate that including undetected offshore evasion alone could reverse the main pattern in Figure 1b, suggesting under-reported income is just as concentrated at the top as reported income.[22]

The failure to audit pass-through businesses owned by individuals subject to random audits is likely to be a significantly larger source of bias than offshore wealth. Pass-through income is about 1/3 of economic income in the top 1%, and estimated rates of under-reporting in pass-through businesses are substantial – the most recent S corporation random audit study found 20% of S corporation income was under-reported without any DCE-type adjustments. This figure falls, realistically, between the Tax Gap figure for sole proprietorships (without DCE) and C corporations. Based on these figures, Guyton et al. (2023*b*) show that the bias from missing pass-through under-reporting is about twice as large as for offshore evasion.[23] Accounting for

[22]Guyton et al. (2023*a*) present sensitivity analysis around the parameters of the exercise we sketched here; virtually all available evidence suggests that offshore evasion is highly concentrated in the top 1% by income, but the taxable rate of return on this wealth and the aggregate extent of concealed offshore wealth are more uncertain (especially over time, discussed further below).

[23]The comparison of the aggregate importance of these two factors is complicated by their overlap: individuals can receive financial capital income via pass-through businesses that own offshore assets, and offshore intermediaries can be used to shelter business income from taxation. The benchmark figures we draw from Guyton et al. (2023*b*) generally attribute the overlap to the offshore component for financial income and to the pass-through business component for business income.

pass-through under-reporting could more than reverse the pattern from Figure 1b, and when we contemplate offshore and pass-through under-reporting together, all signs point toward downward bias.

Are there undetectable forms of under-reported income in the bottom 99% that could offset some of this bias? We concede that it is unclear whether DCE captures all undetected under-reporting in the bottom 99%, but we are skeptical that undetectable under-reporting in the bottom 99% could meaningfully offset the bias coming from undetectable evasion in the top 1%. The DCE procedure already more than doubles under-reporting in sole proprietorships relative to what auditors detect, and an additional (small) adjustment for unreported tip income is already included in both AS and PSZ's estimates.

4.4 Dynamics

The previous discussions were focused on sources of bias when using the most recent waves of random audit data, from about 2006 to 2013, to estimate the distributional properties of unreported income. But how do these factors shape the estimated evolution of inequality over time? Given the immense difficulties involved, we argue that some bias in the level of the top 1% National Income share is probably unavoidable, but a far more important question is whether this bias evolves over time in a fashion we might mistakenly attribute to real changes in income inequality.

Given what we see about the importance of under-reported proprietor income in NIPA over time in Figure 1a, the single most important dynamic source of bias here is probably due to the well-documented rise of pass-through business structures. The top 1% of the income distribution received about 11% of its economic income through pass-through businesses in 1980, and this share grew to 36% in 2019 (for fiscal income only, these shares and the change over time are larger).[24] If pass-through income at the top is consistently under-estimated, the resulting bias, expressed as a share of national income, could therefore grow by a factor greater than 3 from 1980 to present. Growth in pass-through income accounts for almost all of the growth in top 1% national income shares estimated by PSZ.

We can also get some clues about the dynamics of offshore evasion from available evidence. Macro estimates suggest that wealth in offshore tax havens rose substantially from the 1980s to about 2008 when a global crackdown began (Zucman, 2013). Since 2008, the available evidence suggests that offshore evasion is on the decline globally. The findings of two studies in the US (and a few other studies on similar crackdowns internationally) suggest modest gains have been made in the fight against offshore tax evasion (De Simone, Lester and Markle, 2020; Johannesen et al., 2020, 2023). The exact magnitude of any recent decline in offshore evasion is highly uncertain (see Johannesen et al. (2023) for some discussion and illustrative calculations). We note that Auten and Splinter claim that the recent decline in offshore evasion is likely substantial – in our view, the jury is still out on this question – but they do not discuss the likely rise in offshore evasion during the period in which their estimates of top 1% income shares and those of PSZ diverge.

We identify a third dynamic factor could be quantitatively important but whose importance

[24] These figures are drawn from PSZ as reported in Guyton et al. (2023*b*) Figure A7. Similar statistics are reported in Smith et al. (2019) and other work by these authors.

is uncertain with the data available to us: changes to audit procedures in the random audit data over time. The main change involves the transition from the TCMP, which ran the random audit studies between 1963 and 1988, to the NRP, which ran random audit studies in 2001 and afterwards. The transition from TCMP to NRP was motivated in part by concerns about the burden imposed by TCMP audits on taxpayers, so NRP audits were designed to estimate non-compliance across the entire tax return using less intensive procedures (Brown and Mazur, 2003). It is natural to wonder if these changes in audit procedures could create bias that matters differentially at the top of the distribution, where audits are usually more intensive. Indeed, our reading of the literature is that DCE was introduced to accommodate the more limited coverage of noncompliance in NRP compared to TCMP; Slemrod (2007) discusses these changes and recommends caution when comparing NRP to TCMP estimates in the aggregate, let alone at the distributional level. To our knowledge, the only explicit distributional comparisons of data from TCMP and NRP are those in Auten and Langetieg (2023), but because their results are broken down by reported income rather than exam-corrected income, we cannot gain with confidence any insights about how the concentration of detected under-reporting changes over time in various waves of random audit data. A more direct analysis of this question – examining how estimated concentration like Figure 1b varies across waves – using the data from Auten and Langetieg could shed light on this question.

The extent to which some of the factors discussed above contribute to the increase in unreported non-farm proprietor income as a share of National Income from Figure 1a is unclear. One simple question is how much of the increase in Figure 1a is due to partnerships versus sole proprietorships.[25] This cannot be inferred from public tax gap statistics because estimated partnership under-reporting is not reported separately from S corporation under-reporting. Based on the facts reviewed above, however, we find it plausible that the increase is entirely driven by changes to random audit procedures, DCE methods, and the increase in the importance of partnership income over time. The first two suggest proprietor under-reporting may not have actually grown in importance at all, while the third suggests that if it did, the change should load mainly onto the top of the distribution, where partnership income is concentrated.

In criticizing AS, PSZ describe what they see as unrealistic features of AS' estimates (Piketty, Saez and Zucman, 2023). They take a more macro perspective, showing the AS aggregate estimates imply that untaxed business income is much less concentrated than taxed business income, to a growing extent over time. PSZ argue this is unrealistic in light of empirical data (mainly data on reported incomes and prior estimates of wealth inequality). Our work here isolates one mechanism by which the patterns PSZ found unrealistic come about. By imposing that under-reported income belongs mainly to the bottom 99%, AS' micro-simulations impose that over time, the growth of unreported pass-through income is less concentrated than the growth of reported pass-through income. Moreover, the structure imposed on distributional estimates is implicit in micro-simulations (while the micro structure becomes more explicit). Whether this feature of the AS estimates is credible depends on the credibility of the assumption that random audit data identify the correct concentration, consistently over time.

[25] We have asked experts at both the BLS and IRS for clarity on this question without any luck.

5 Alternative Approaches

In this Section, we critically review the approach of PSZ to under-reported income and consider some alternatives.

5.1 Methods in PSZ

At the individual level, PSZ assume that unreported income is proportional to reported positive income. They scale the reported positive income of each income type j by individual i in year t by a constant multiplier $m_{jt} = \frac{Y_{jt}^{NIPA}}{\sum_i \max\{y_{ijt},0\}}.$, i.e. income is re-scaled at the individual level to $m_{jt} * \max\{y_{ijt}, 0\}$. Their approach does not separately account for the different reasons NIPA totals may diverge from their fiscal income analogues (bonus depreciation, different data sources, under-reporting, etc.).

At the individual level, one obviously unrealistic feature of this allocation, where unreported income is concerned, is that PSZ do not assign any under-reporting to individuals with reported business losses. AS argue that this feature of PSZ's approach generates upward bias in their estimates of the top 1% share. In random audits, a substantial portion of under-reporting of business income is detected for those with reported losses. Due to re-ranking, an allocation proportional to positive income causes those who have already reported a high income to be allocated even more income and thus to move further up in the income distribution; AS argue that this introduces upward bias into the estimates. We fully agree that at the micro level, this is an unrealistic allocation of under-reported business income. In fact, the microsimulations in both of these papers and in IRS Tax Gap studies all entail obviously unrealistic features.[26] The relevant question in our view is how potentially unrealistic features of the micro-simulation could introduce bias in the aggregate, after re-ranking.

At the aggregate level, the bias from using positive incomes to allocate under-reporting is *ambiguously signed.* To see why, it is useful to consider an extreme scenario in which everyone who over-claims business losses winds up in the top 1% by corrected income. In that case, allocating no under-reporting to those with reported losses would bias the estimated top 1% share downwards because 100% of the misallocated income belongs in the top 1%. In general, the question is whether the concentration of reported and unreported income that results from PSZ's specification exceeds or is exceeded by the true concentration at the top of the true income distribution. Regarding the use of random audit data to address under-reporting for those with reported losses in particular, we note that many over-claimed losses involve partnerships, especially at the top of the income distribution. In our view, without comprehensive random audit data on partnerships, it is impossible to be confident about the extent to which PSZ's approach is biased. For that reason, we do not agree with AS' characterization of this feature of PSZ's estimates as an obvious flaw that they have corrected, but rather view this as a question about the implications of differing assumptions about the distribution of unreported income.

In any case, the divergence between AS and PSZ derives not only from the fact that AS

[26]For example, AS assume that all undetected under-reporting belongs to individuals with nonzero detected under-reporting. Individuals with no detected under-reporting are never assigned undetected under-reporting. Older micro-simulations used by IRS in its Tax Gap studies share this property, but the micro-simulations used in Tax Gap studies since 2019 makes the exact opposite assumption (which is also unrealistic): they allocate all undetected under-reporting along the extensive margin of detection (see Guyton et al., 2023*a*, for more discussion).

allocate some under-reporting to those with reported losses, but also from differences in assumptions about whether the concentration in the top 1% of the distribution is better approximated by the concentration of audit-detected mis-reporting or by the concentration of reported income. Our next exercise disambiguates the importance of these two factors.

5.2 Our Proposed Method and Comparisons

Here we propose an alternative method–building on some ideas in Guyton et al. (2023*a*) and Guyton et al. (2023*b*)–that captures the main rationale for PSZ's allocation but is not subject to the criticism from AS about potential excess re-ranking. We make use of publicly available Distributional National Accounts micro-data. There are annual files for each year (starting in 1962, with gaps in 1963 and 1965), where each row is a synthetic individual and each column is a matching income or demographic value. For more information, see the DINA page here.

First, we introduce an accounting identity decomposing the top 1% share of national income, denoted Y_{1t}, into components by type of income and whether the income was reported. Letting M_t be the share of national income that is misreported income, by construction we have

$$Y_{1t} = (1 - M_t) \sum_j S^R_{jt} R_{1jt} + M_t \sum_j S^M_{jt} X_{1jt}. \tag{3}$$

Let y_{ijt} and x_{ijt} denote the reported and unreported income, respectively, of taxpayer i in year t of type j. The shares $S^R_{jt} = \frac{\sum_i y_{ijt}}{\sum_i \sum_j y_{ijt}}$ and $S^M_{jt} = \frac{\sum_i x_{ijt}}{\sum_i \sum_j x_{ijt}}$ capture the importance of income type j in total reported and unreported income, respectively. We also define

$$R_{1jt} \equiv \frac{\sum_{i \in \text{top 1\% by national income in yr } t} y_{ijt}}{\sum_i y_{ijt}},$$

and X_{1jt} is defined analogously replacing y_{ijt} with x_{ijt}. In words R_{1jt} and X_{1jt} are the share of reported and misreported income of type j, respectively, that belong to the top 1% *of the true/national income distribution.*[27]

Our proposed method leverages the accounting identity above, together with the following two assumptions.

Assumption 1: Irrelevance of Re-ranking/Negatives for Reported Income Distribution.

$$R_{1jt} = \frac{\sum_{i \in \text{Top 1\% by reported inc in yr } t} \max\{y_{ijt}, 0\}}{\sum_i \max\{y_{ijt}, 0\}}.$$

Assumption 2: Distributional Neutrality by Type of Income. For every income type j in every year t, $X_{1jt} = R_{1jt}$.

Assumption 1 allows us to use the observed concentration of reported positive income when ranking tax units by reported income as a proxy for the concentration of reported income when ranking by true income. The assumption mainly requires that the concentration of reported income is negligibly affected by the allocation of under-reporting and re-ranking by true income; we validate this notion empirically using aggregates from NRP data below. One could

[27] Note that we define R_{1jt} slightly differently here compared to when we introduced Figure 1d/equation (2) in the Introduction; any confusion this causes should be rectified by Assumption 1, which we maintained implicitly in the Introduction.

obviously modify the assumption to address negative incomes in different ways. We explore sensitivity to how we treat negatives below.

More importantly, Assumption 2 captures the main rationale we find in PSZ's defense of their work, which we find persuasive: if reported income of some type becomes much more unequally distributed, then in the absence of reliable data, it is natural to assume that unreported income of the same type also becomes more unequally distributed. A necessary and sufficient condition for Assumption 2 to hold is if the fraction of each type of income that is misreported is the same in the top 1% by true income as in the full population. Formally, it is straightforward to show that

$$\forall j, t, \ \frac{\sum_{i \in \text{Top 1\%}} x_{ijt}}{\sum_{i \in \text{Top 1\%}} y_{ijt} + x_{ijt}} = \frac{\sum_i x_{ijt}}{\sum_i y_{ijt} + x_{ijt}} \iff \forall j, t, \ X_{1jt} = R_{1jt}.$$

If we replace Assumption 2 with the assumption that misreporting is distributed like audit-detected misreporting – $X_{1jt} = E_{1t}$ for each j – then we obtain the figures based on AS in Equation 2 and Figure 1d. By similar logic to the above, the sign of the bias in our benchmark depends on the differences between the true parameter, X_{1jt}, and R_{1jt}, for each type of income; which types of income matter most for bias is governed by S^M_{jt}. Relatedly, because we allocate under-reporting in proportion to positive income in the top 1% by reported income, this specification is equivalent at the aggregate level to using the PSZ micro allocation without re-ranking after allocating unreported income. We do not claim this benchmark is clearly the best approach to calibrating a distribution of unreported income, but at minimum it provides a useful way to understand the sources of disagreement and uncertainty here.

We implement this approach, decomposing income into four types: wages, rental income, s-corp income, and business income (defined here to include farm and non-farm proprietor income). Figure 2 illustrates the difference between the concentration of reported versus corrected income using the PSZ micro allocation, our benchmark, and the AS assumptions (reproduced from Figure 1d). Our benchmark agrees with PSZ about the qualitative level and dynamic effect on the top 1% income share. The largest adjustments for under-reporting are due to pass-through business income, which is more concentrated at the top than overall income in reported income data, so the level increases somewhat. More importantly, in contrast to AS, the top 1% share increases over time in both our benchmark and in PSZ. As we confirm shortly, this increase in the top 1% share of income is driven by the same force in both series: the rise of pass-through business forms at the top and the resulting increase in the concentration of pass-through business income. However, the PSZ approach features excess re-ranking of income into the top 1% compared to our benchmark, to a somewhat growing extent over time. As a result, increase in the top 1% share of income between 1962 and 2014 is 0.63 percentage points smaller in our series than in our simulation of PSZ. This difference arguably reflects the isolation and correction of the specific feature of PSZ's approach that AS find unrealistic – excess re-ranking due to those with high positive income being allocated high under-reporting at the micro level – but it is rooted in the same overall rationale that PSZ present in defense of their approach, as made explicit in Assumption 2.

As a brief note, with this figure, we can reasonably replicate the 1.6 percentage point difference in the growth of the share of income held by the top 1 percent between AS and PSZ by comparing the PSZ Micro and AS Macro approaches. Our macro-approach based on AS implied that their methodology decreased the top 1 percent share by 0.46 percentage points between 1962 and 2014, while the micro-approach following PSZ finds a matching increase of 1.12 percentage points, for a difference of 1.58 percentage points. There are several reasons that this comparison likely misses some components of the 1.6 percent gap calculated by AS, including our use of a constant E_1 in our AS macro approach and the influence of other components of NIPA that might complicate re-ranking effects. However, by comparing the macro-estimates we produce, we can isolate the effect of the underlying assumptions made by both sets of authors: the 0.46 percentage point decrease from AS versus a 0.52 percentage point increase from our benchmark based on distributional neutrality. Together, the differences in assumptions explain 0.98 percentage points of the gap, or approximately 61 percent. We can therefore be reasonably confident that the gap is driven primarily by different conceptual assumptions regarding misreporting rather than differences in micro-simulation methodology.

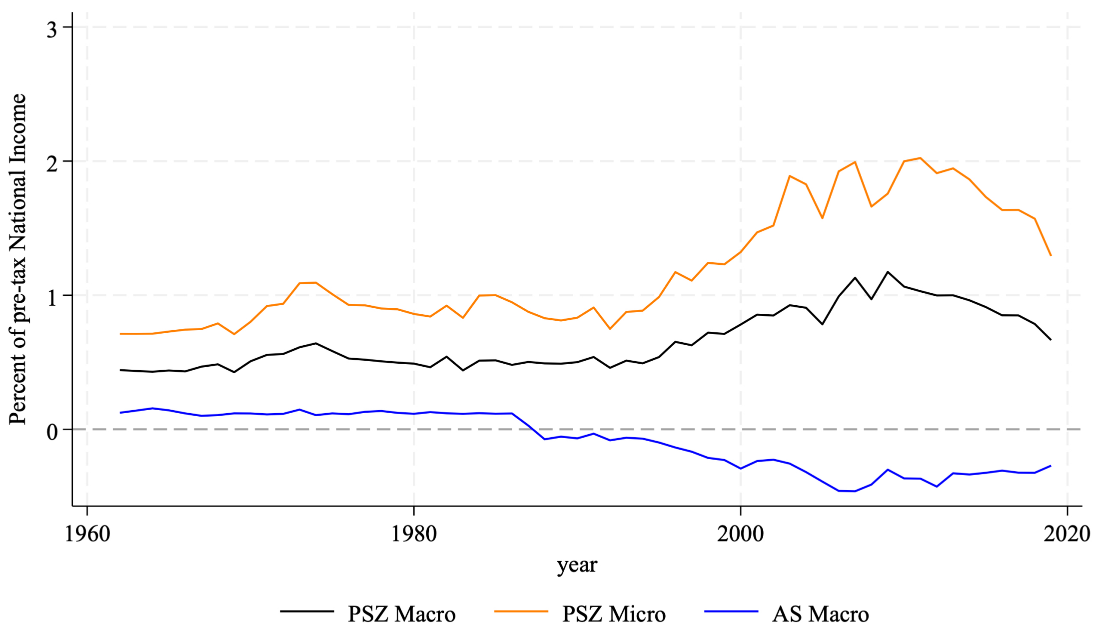
Our benchmark suggests that accounting for unreported income increases the top 1% income share by 0.52 percentage points between 1962 and 2014. In Figure 3, we explain why this occurs and do some sensitivity/validation exercises. The black series in Panels (a)-(c) of Figure 3 is the difference between Y_{1t} and R_{1t}, or the change in the share of income held by the top 1 percent by allocating wage and business misreporting in NIPA proportional to each type of reported income. The other series these panels illustrate what we would have found under different assumptions.

Figure 3a illustrates the role of changes in misreporting as a share of income (from Figure 3a) and changes in the concentration of each type of income. To do so, we hold these quantities fixed at their 1962 levels and recompute the estimates. When we hold the concentration of reported income of each type of income fixed, the dynamic effect of adding under-reporting changes from an increase to a decrease in the top income share, and when we hold the importance of unreported income as a share of national income fixed, almost all of the dynamic increase vanishes. Figure 3a therefore reveals the dynamic increase in the top 1% income share in our benchmark (and in PSZ) derives from an increase in the reported-income concentration of certain types of income, together with increases in the importance in NIPA of unreported income of these same types of income. All evidence points to pass-through business income being the main type of income that drives all this (Figure 1a, the rise of pass-through at the top, etc.). We confirm this in Figure 3b, in which we hold the distribution of non-wage income alone fixed throughout the period and re-compute the top income share. In other words, for $j =$business income, both S^X_{jt} and R_{1jt} increase significantly over time.

In Figure 3c, we explore some alternative ways of modifying how we accommodate negative incomes in Assumption 1: leaving negative incomes as is and assigning negative misreporting in proportion to them, or using absolute values. What we find suggests that including negatives as-is would have increased the top 1 percent share by 0.84 percentage points between 1962 and 2014, verses 0.52 percentage points in our baseline. It is worth noting, however, that there are several periods of extreme noise when including negatives. Treating negatives as absolute values would marginally reduce the top one percent share. As discussed above, it is difficult to discern the correct way to handle negatives without a model or better audit data, but we conclude from this analysis that alternative ways of implementing the overall idea of distributional neutrality by type of income in the presence of negative values for income will not change the qualitative results. [28]

Finally, Figure 3d tests the other main requirement of Assumption 1, which is that the concentration of reported income at the top is unaffected by re-ranking by true income. We provide empirical validation for this assumption using random audit estimates from Guyton et al. (2023*b*) (without DCE) in Figure 3d. The figure illustrates how the concentration of reported and unreported income depends on re-ranking by exam-corrected versus reported incomes. We find that the distribution of unreported income is very different between the two, as expected, but the distribution of reported income is virtually identical whether we rank by reported or exam-corrected income, so much so that it is difficult to see the difference between the two estimates because the lines are so close together.

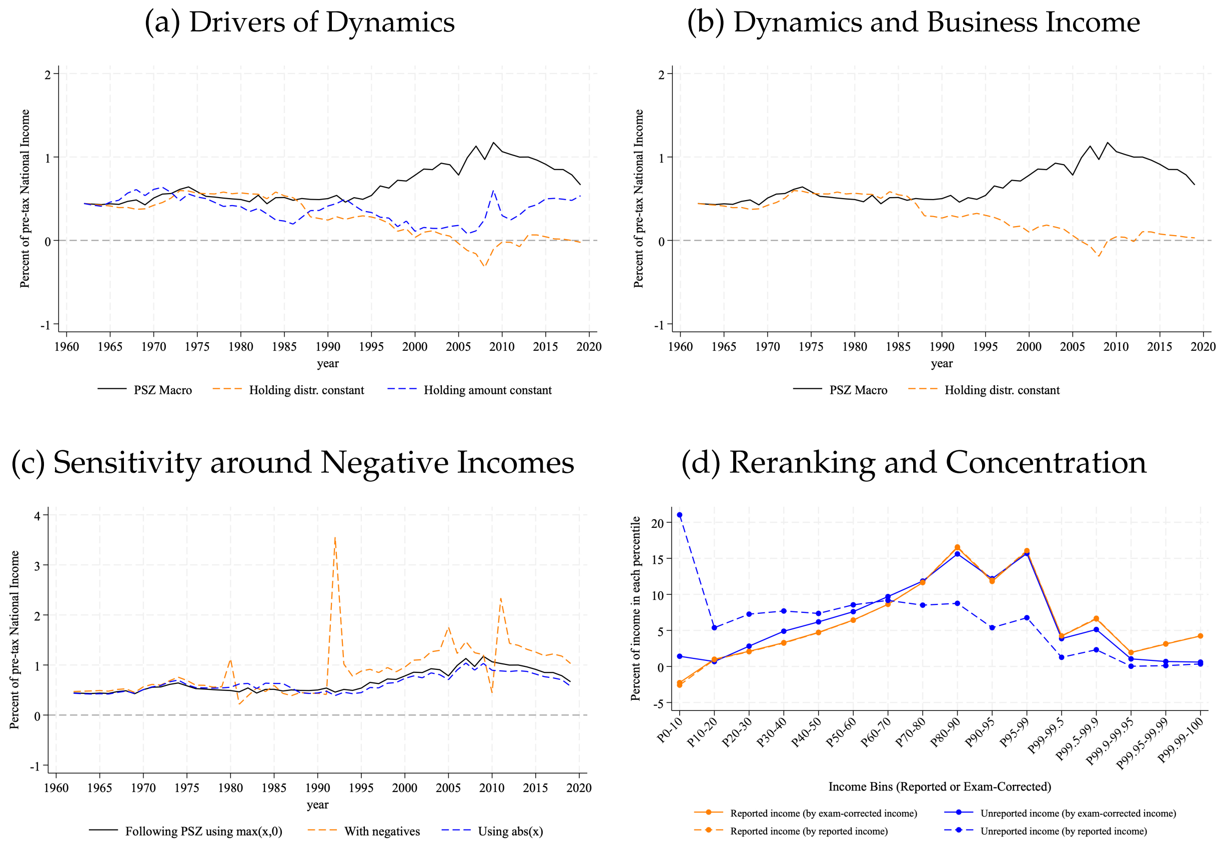
FIGURE 2: UNDER-REPORTING AND DYNAMICS OF THE TOP 1% SHARE OF PRE-TAX NATIONAL INCOME: COMPARISON OF ALTERNATIVE METHODS



Notes: This figure illustrates the effect of adding misreported income to national income under the assumption that national income sans misreported income and reported fiscal income exhibit similar concentration. As in Figure 1d, we plot the difference between the top 1% national income share with and without including misreported income. We compare three different methods: a micro simulation based on PSZ with misreporting allocated proportionally to positive income by type of income (in orange), the macro approach that mimics AS (identical to Figure 1d) with misreporting allocated proportionally based on the share of misreporting in random audit data (in blue), and our preferred approach, a macro approach with misreporting proportionally to reported positive income by type of income (in black). Our preferred approach generates qualitatively similar estimates to the method of PSZ, with a slightly smaller increase in the top income share over time.

[28]In a comment on an earlier draft of Guyton et al. (2023*b*), Auten and Splinter (2021) criticize a similar macro approach to allocation of unreported pass-through income at the distributional level. The criticism mirrors their criticism of PSZ and appears to confuse the micro-level proportional-allocation-before-re-ranking approach of PSZ to the macro approach based on equation (3) and Assumptions 1 and 2. This is discussed further in Reck, Risch and Zucman (2021); it may not have been clear in the initial draft of Guyton et al. (2023*b*) what method was being used.

FIGURE 3: UNDER-REPORTING AND DYNAMICS OF THE TOP 1% SHARE OF PRE-TAX NATIONAL INCOME: UNDERSTANDING OUR BENCHMARK



Notes: The solid black line in Panels (a)-(c) plots the difference in the top 1% share between our benchmark and reported income. Our benchmark is a macro allocation (see eq. (3) plus assumptions 1 and 2), with misreporting distributed proportionally to the share of positive reported income (separately calculated for wage and business income) claimed by the top 1 percent. The blue dashed line in Panel (a) holds constant the total amount of misreported income at 1962 levels, but allows the distribution to vary over time; the orange dashed line holds constant the distribution of misreported income at 1962 levels, but allows the amount of misreporting to vary over time. In Panel (b), we perform an analogous exercise and only hold the distribution of business income constant. From these results, we infer that the growth in inequality implied by our benchmark is driven primarily by increases in the concentration of reported business income relative to all types of reported income. Panel (c) explores sensitivity of our benchmark to different ways of handling negative income observations. Rather than using an allocation proportional to positive business income, the orange and blue dashed lines either leave negative profits unchanged in the calculation of M_{1t} (orange), or use the absolute value of negative reported business income values (blue). Panel (d) plots shares of reported (orange) and unreported (blue) income through the distribution, using NRP data for tax years 2006–2013 from (Guyton et al., 2023*b*). We include only audit-detected under-reporting (no DCE). The solid lines represent income measures sorted by exam-corrected income percentiles, while the dashed lines represent income measures sorted by reported-income percentiles. We observe that the concentration of reported income is unaffected by re-ranking, which suggests that our allocation does not create mechanical bias by imposing unrealistic structure on reranking.

In the end, we argue that we do not have sufficient empirical data to make confident claims about which of these approaches is best. Our subjective view is that in the absence of clear evidence, the type of distributional neutrality we require in our benchmark is desirable. PSZ and others argue that we have high-quality data on reported income that suggests inequality is on the rise, and too little data on unreported income with which to challenge this assumption. We find this argument persuasive. However, in our view, explicitly ensuring that we respect the concentration of reported income of each type in a macro approach is a better way to build on this argument than the micro allocation used by PSZ, due to the excess re-ranking discussed above. Nevertheless, we emphasize that if under-reporting is more concentrated at the top than reported (positive) income, to an increasing extent over time, PSZ's approach is likely less biased than our benchmark. The preferred estimates in Guyton et al. (2023*b*) do indeed suggest that under-reporting is a larger share of income in the top 1% than in the bottom 99%. This would suggest that the level of the top 1% income share in our benchmark is too low in 2006–2013, but it is unclear what this means dynamically. Meanwhile if the concentration in random audit data more closely captures the true concentration of unreported income (i.e. if the sources of bias described above are unimportant), the approach by AS could be closer to reality.

Some limitations of our benchmark seem noteworthy. Because we have focused here solely on unreported income, we note that way in which one accounts for other differences between fiscal and national income could interact with the estimated concentration of unreported income. We do not account for the importance of any of this. Doing so would be straightforward, but it requires confronting additional sources of uncertainty and disagreement between AS and PSZ. Additionally, a useful practical feature of PSZ and AS is that they create a micro-level dataset in which all components of national income are allocated at the micro level. Our macro specification does not entail such a micro-level allocation of unreported income. This reflects the pros and cons of our approach versus micro-simulation approaches more broadly: our approach makes the structure imposed on the aggregate distribution of income more explicit, while the structure imposed on micro-level allocation of income and re-ranking becomes implicit. Nevertheless, it should be feasible to find a micro allocation of unreported income that is consistent with our Assumptions 1 and 2. This would require some statistical modelling that we do not work through here.

Finally, for both our benchmark and PSZ, we observe that relying on reported incomes to allocate unreported components of income *reduces the amount of new information revealed by the estimates, compared to PS.* For example, consider the fact that PSZ estimate a similar increase in the top 1% income share over time to PS. If PSZ were able to observe all components of national income at the micro level the way we can observe reported fiscal income on tax returns, the finding that the increase in the top 1% income share is similar between PSZ and PS could be treated as *confirming* the findings of PS about rising inequality. If instead, all components of National Income were allocated in proportion to reported incomes on tax returns, then it would be mechanically ensured that the increase in inequality is similar between National and Fiscal income concepts. In this case, we should not update our beliefs about rising inequality at all from comparing PSZ to PS. The correct amount of belief updating clearly falls somewhere between these two extremes. Additional work on the sensitivity of the PSZ estimates to different assumptions and empirical evidence related to these assumptions would help us to further resolve our uncertainty about the distribution of National Income and its dynamic evolution.

6 Conclusion

In this comment, we took a deep dive into the question of the distributional allocation of unreported income and its importance for the dynamics of top income inequality. We find, after close examination of their methods, that AS' results rest on an implicit assumption that all misreporting included in National Income is distributed like audit-detected misreporting, even though the former exceeds the latter by a factor of 3. We caution against reversing the consensus that the top 1% share of income has risen sharply in the United States in the last several decades on the basis of this assumption. Owing to undetected under-reporting and uncertainty around DCE, estimates of the level and concentration of under-reporting based solely on random audit data should be interpreted with extreme caution. These uncertainties grow over time as statistical methods, audit procedures, and real features of the economy all change, so that we cannot make confident claims about the dynamics of inequality using this type of data. Apart from the under-reporting component of income, dynamic estimates suggest that inequality is rising and we see no credible evidence that accounting for under-reporting should reverse this trend.

Meanwhile, the approach of PSZ is broadly consistent with the notion that if reported business income has grown in importance and concentration at the top, then so has unreported business income. We propose a simple benchmark method for accounting for unreported income in the aggregate that is rooted in this same notion, via an assumption of distributional neutrality of each type of unreported income. Relative to this benchmark, in which the top 1% income share grows by about 0.52 percentage points over time, the PSZ micro simulations include a re-ranking effect that causes the top 1% income share to grow by an additional 0.6 percentage points. We do not claim our benchmark is clearly superior to the approach of PSZ and AS, but it has the virtues of simplicity and distributional neutrality around unobserved components of income.

Virtually all direct empirical analysis we have seen points in the same direction: pre-tax income inequality has risen dramatically in the United States since the 1980s, especially at the very top of the distribution. This includes not only analysis of tax return data, but also research on on CEO pay (Murphy, 2013; Piketty, 2013), within-firm wage inequality (Zwysen, 2022; Wallskog et al., 2024), and rising wage inequality between firms (Song et al., 2019). Rising inequality over the last few decades is also essential to making sense of other real changes in the US economy, such as the relationship between technological change and inequality (Autor, Katz and Kearney, 2008; Acemoglu and Autor, 2011) and between debt and inequality (Bartscher et al., 2020; Mian et al., 2021). That inequality was rising significantly was already a widespread view when PS published their study in 2003; the new information in their study involved the extent to which this was driven by incomes at the very top of the distribution, e.g. the top 1% as opposed to the top 20%. They were able to examine this because tax return data contain uncommonly comprehensive coverage of income at the very top of the distribution. To be sure, this coverage is imperfect and does not include all components of National Income, including under-reported income. But the evidence we reviewed reveals that the coverage of unreported income at the very top of the distribution in random audit data is extremely limited, and we argue these data are insufficient to credibly overturn prior findings of substantial increases in inequality in the last few decades.

Given the available data, there is a lot of uncertainty in estimates of Distributional National Accounts (DINA). This has led to some challenging debates, but it also creates opportunities for researchers to contribute to this discussion. One informative exercise would be to unpack the model uncertainty in DINA estimates from a formal or informal Bayesian perspective, starting from a set of estimates like PSZ and illustrating not just how one alternative approach plays out, as AS attempted to do, but what the main sources of model uncertainty even are, in an empirically grounded way. More ambitiously, there is a deep econometric problem at the heart of this question. If we observe the distribution of reported income z, and we want to know about the distribution of national income $y = z + e$, and we are mainly interested in the top 1% income share, what are the "sufficient statistics" about the joint distribution of z and e that we ought to estimate in order to characterize how the top 1% share of y is different from that of z? Under what primitive assumptions would a macro approach like our benchmark or a micro simulation like that of AS or PSZ be unbiased? How might knowing the parametric structure of these distributions (e.g. that they have Pareto tails) help us to simplify the problem? From conversations with some experts, our understanding is that these econometric questions are not well-understood in general. Answering them would help us build a consensus on the important question of how much income inequality in the United States has really risen in the last four decades.

References

Acemoglu, Daron, and David Autor. 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." In *Handbook of Labor Economics*. Vol. 4, 1043–1171. Elsevier.

Alstadsæter, Annette, Martin Jacob, Wojciech Kopczuk, and Kjetil Telle. 2023. "Accounting for Business Income in Measuring Top Income Shares: Integrated Accrual Approach Using Individual and Firm Data from Norway." National Bureau of Economic Research w22888, Cambridge, MA.

Alstadsæter, Annette, Niels Johannesen, and Gabriel Zucman. 2018. "Who owns the wealth in tax havens? Macro evidence and implications for global inequality." *Journal of Public Economics*, 162: 89–100.

Auten, Gerald, and David Splinter. 2021. "Comment: Tax Evasion at the Top of the Income Distribution: Theory and Evidence." *Comment*.

Auten, Gerald, and David Splinter. 2023. "Income Inequality in the United States: Using Tax Data to Measure Long-Term Trends." *Journal of Political Economy*, 728741.

Auten, Gerald, and Patrick Langetieg. 2023. "The Distribution of Unreported Income: What We Can Learn from the NRP."

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. 2008. "Trends in U.S. Wage Inequality: Revising the Revisionists." *Review of Economics and Statistics*, 90(2): 300–323.

Bartscher, Alina K., Moritz Kuhn, Moritz Schularick, and Ulrike I. Steins. 2020. "Modigliani Meets Minsky: Inequality, Debt, and Financial Fragility in America, 1950-2016." Federal Reserve Bank of New York Staff Report 924, New York.

Black, Emily, Ryan Hess, Rebecca Lester, Jacob Goldin, Daniel E Ho, Mansheej Paul, and Annette Portz. 2023. "The Spiderweb of Partnership Tax Structures."

Black, Theodore, Kim Bloomquist, Edward Emblom, Andrew Johns, Alan Plumley, and Esmeralda Stuk. 2012. "Federal Tax Compliance Research: Tax Year 2006 Tax Gap Estimation." *IRS Research, Analysis & Statistics Working Paper*.

Brown, Robert E, and Mark J Mazur. 2003. "The National Research Program: Measuring Taxpayer Compliance Comprehensively." *Kansas Law Review*, 51(5): 1255–1272.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014*a*. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review*, 104(9): 2593–2632.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014*b*. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review*, 104(9): 2633–2679.

Cooper, Michael, John McClelland, James Pearce, Richard Prisinzano, Joseph Sullivan, Danny Yagan, Owen Zidar, and Eric Zwick. 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, Lisa, Rebecca Lester, and Kevin Markle. 2020. "Transparency and Tax Evasion: Evidence from the Foreign Account Tax Compliance Act (FATCA)." *Journal of Accounting Research*, 58(1): 105–153.

Erard, Brian, and Jonathan Feinstein. 2011. "The Individual Income Reporting Gap: What We See and What We Don't." Internal Revenue Service Publication 1500 (rev 4-2012), Washington D.C.

Erard, Brian, and Jonathan S. Feinstein. 2010. "Econometric Models for Multi-Stage Audit Processes: An Application to the IRS National Research Program." In *Developing Alternative Frameworks for Explaining Tax Compliance*. . 1st ed., , ed. James Alm, Jorge Martinez-Vazquez and Benno Torgler. London:Taylor & Francis Group.

Feinstein, Jonathan S. 1990. "Detection Controlled Estimation." *The Journal of Law and Economics*, 33(1): 233–276.

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

Gordon, Roger H., and Joel Slemrod. 2002. "Are "Real" Responses to Taxes Simply Income Shifting Between Corporate and Personal Tax Bases?" In *Does Atlas Shrug? The Economic Consequences of Taxing the Rich..* , ed. Joel Slemrod. Cambridge, MA:Harvard University Press.

Guyton, John, Patrick Langetieg, Daniel Reck, Max Risch, and Gabriel Zucman. 2023*a*. "The Distribution of Undetected Under-Reported Income Identified by Detection-Controlled Estimation Methods." (Availible in the appendix of Guyton et al. (2023b)).

Guyton, John, Patrick Langetieg, Daniel Reck, Max Risch, and Gabriel Zucman. 2023*b*. "Tax Evasion at the Top of the Income Distribution: Theory and Evidence." National Bureau of Economic Research Working Paper 28542, Cambridge, MA.

Hemel, Daniel J., Janet Holtzblatt, and Steve Rosenthal. 2021. "The Tax Gap's Many Shades of Gray." *SSRN Electronic Journal*.

Jason Debacker, Bradley Heim, Anh Tran, and Alexander Yuskavage. 2020. "Tax Noncompliance and Measures of Income Inequality." *Tax Notes*.

Johannesen, Niels, Daniel Reck, Max Risch, Joel Slemrod, John Guyton, and Patrick Langetieg. 2023. "The Offshore World According to FATCA: New Evidence on the Foreign Wealth of U.S. Households." In *Tax Policy and the Economy, Volume 38*. Vol. 38, , ed. Robert A. Moffitt. University of Chicago Press.

Johannesen, Niels, Patrick Langetieg, Daniel Reck, Max Risch, and Joel Slemrod. 2020. "Taxing Hidden Wealth: The Consequences of US Enforcement Initiatives on Evasive Foreign Accounts." *American Economic Journal: Economic Policy*, 12(3): 312–346.

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

Johnson, Barry W., and Peter J. Rose. 2019. "Federal Tax Compliance Research: Tax Gap Estimates for Tax Years 2011–2013." Internal Revenue Service, Research, Applied Analytics & Statistics 1415, Washington D.C.

Love, Michael. 2021. "Where in the World Does Partnership Income Go? Evidence of a Growing Use of Tax Havens." *SSRN Electronic Journal*.

Mazur, Mark J., and Alan H. Plumley. 2007. "Understanding the Tax Gap."

Mian, Atif, Ludwig Straub, Amir Sufi, and Chicago Booth. 2021. "The Saving Glut of the Rich." *Working Paper*.

Murphy, Kevin J. 2013. "Chapter 4 - Executive Compensation: Where We Are, and How We Got There." In . Vol. 2 of *Handbook of the Economics of Finance*, , ed. George M. Constantinides, Milton Harris and Rene M. Stulz, 211–356. Elsevier. ISSN: 1574-0102.

Piketty, Thomas. 2013. *Capital in the Twenty-First Century.* Cambridge, MA:Harvard University Press.

Piketty, Thomas, and Emmanuel Saez. 2003. "Income Inequality in the United States, 1913–1998." *Quarterly Journal of Economics*, 118(1): 1–41.

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman. 2018. "Distributional National Accounts: Methods and Estimates for the United State." *The Quarterly Journal of Economics*, 133(2): 553–609.

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman. 2023. "Comment on Auten and Splinter (2023)." *Comment*.

Reck, Daniel, Max Risch, and Gabriel Zucman. 2021. "Response to a Comment by Auten and Splinter on "Tax Evasion at the Top of the Income Distribution: Theory and Evidence"." *Comment*.

Slemrod, Joel. 2007. "Cheating Ourselves: The Economics of Tax Evasion." *Journal of Economic Perspectives*, 21(1): 25–48.

Smith, Matthew, Danny Yagan, Owen Zidar, and Eric Zwick. 2019. "Capitalists in the Twenty-First Century." *Quarterly Journal of Economics*, 134(4): 1675–1745.

Smith, Matthew, Owen Zidar, and Eric Zwick. 2022. "Top Wealth in America: New Estimates Under Heterogeneous Returns." *The Quarterly Journal of Economics*, 138(1): 515–573.

Song, Jae, David J Price, Fatih Guvenen, Nicholas Bloom, and Till Von Wachter. 2019. "Firming Up Inequality*." *The Quarterly Journal of Economics*, 134(1): 1–50.

Splinter, David. 2023*a*. "Reply: Measuring Income Inequality: A Primer on the Debate." *Comment*.

Splinter, David. 2023*b*. "Short Reply: Comment on Auten and Splinter (2023)." *Comment*.

Wallskog, Melanie, Nicholas Bloom, Scott W. Ohlmacher, and Cristina Tello-Trillo. 2024. "Within-Firm Pay Inequality and Productivity." *NBER Working Paper No. 32240*.

Werfel, Danny, Barry W Johnson, Victoria Bryant, and Wayne K Kei. 2024. "Data Book, 2023." Internal Revenue Service Publication 55-B, Washington D.C.

Zucman, Gabriel. 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.

Zwick, Eric, and James Mahon. 2017. "Tax Policy and Heterogeneous Investment Behavior." *American Economic Review*, 107(1): 217–248.

Zwysen, Wouter. 2022. "Wage Inequality Within and Between Firms: Macroeconomic and Institutional Drivers in Europe." *SSRN Electronic Journal*.

A Additional details on Auten and Splinter's methodology

Following PS, AS begins with a set of annual stratified random samples of tax returns, from which the distribution of income in tax data (referred to as fiscal income) is readily observable at the tax unit level. From this starting point, they employ a process of adjustments, additions, and subtractions to produce a measure of income for each household that is consistent with the income concepts reported in NIPA. The whole process employed by AS can be broken down into five stages, detailed in AS's appendix table B2.

Step 1: Improving fiscal income. This step involves adjustments to the IRS samples that can be completed using information from the IRS. Some are adjustments to the sample construction to account for the difference between tax units and households[29] (e.g. removing filers under 20 or who are dependents). Others add or remove types of income to and from Adjusted Gross Income (AGI) to bring fiscal income more in line with the BEA income concepts (e.g. adding tax exempt interest and excluded dividends, removing capital gains).

Step 2: Expansions to pre-tax income. Several components of pre-tax income are not directly observed on tax returns. The adjustments in this stage aim to allocate income from such components to individual tax returns based on related information reported on tax returns (for example, allocating imputed rent in proportion to deducted real-estate taxes). This section includes the allocation of under-reported income and excess depreciation/LEBI, which is the focus of this comment.

Step 3: Adding transfers to pre-tax income. Adjustments in this stage allocate government transfers, such as social security benefits, unemployment insurance benefits, and the value of health insurance, to tax units based on either reported benefits or demographic characteristics.

Step 4: Calculating after-tax income. In this stage, the value of federal, state, and local taxes are subtracted from the income of tax units.

Step 5: Adding value from government sector to after-tax income. Finally, the value of government deficits and surpluses, as well as government consumption, is added to individual income.

We are interested in the expansion to pre-tax income in stage 2 from the inclusion of misreported income, which is income that should appear on a tax return but is not present. This can be due either to intentional evasion or to misunderstanding, ambiguities in tax rules, or other less nefarious causes. In determining how to account for this gap in the calculation of the distribution of income, there are two critical questions: how big is the gap, and how should the missing dollars be distributed among households?

AS start by calculating the gap between NIPA- and tax-reported total income for each of five different types of income: wages and salaries, rental income, S-corporation income, farm

[29] Tax units are groups of individuals who report their income on the same tax return, or who would have done so if they are nonfilers. Dependents living in separate households can be part of the same tax unit, while unmarried cohabiting adults can be in the same household and different tax units. We do not engage here with the question of accounting for the difference between tax units and households, which is the source of a quantitatively minor disagreement between the estimates.

income, and nonfarm proprietor income. For income type j in tax year t, the gap is calculated as:

$$G_{jt} = Y_{jt}^{NIPA} - \sum_i y_{ijt},$$

where Y_{jt}^{NIPA} is the NIPA total in year t for income type j, and $\sum_i y_{ijt}$ is the total income amount observed on tax returns (i) for the same income type and year, after adjustments made in stage 1 or 2.[30] With one exception we describe shortly, their approach assumes that these differences are entirely driven by misreporting, and that the gap between NIPA and tax-reported income is an accurate estimate of under-reporting.

Two complicating factors are worth noting. First, NIPA does not rely on tax data to estimate total income from rents, S corporations, or farms. Second, NIPA does include explicit adjustments for under-reported income for wages and salary income and non-farm proprietor income, using figures it obtains from the IRS based on official Tax Gap statistics (including DCE).[31] However, for non-farm proprietor income, the overall gap G_t is larger than the explicit adjustment for unreported income due to the additional adjustment NIPA makes for excess depreciation/capital consumption.

For most of the income groups (wages, rents, s-corporations, and farm income), the gap G_{jt} is assumed to be entirely composed of misreporting (M_{jt}). To address the second complication above, for non-farm proprietor income ($j = nfp$), the gap is broken down into three categories: misreporting, (M_{jt}^{NIPA}), adjustments for excess depreciation (D_{jt}^{NIPA}), both obtained directly from NIPA, and a residual factor. In other words:

$$G_{jt} = \begin{cases} M_{jt} & \text{if } j \neq nfp \\ M_{jt}^{NIPA} + D_{jt}^{NIPA} + R_{jt} & \text{if } j = nfp \end{cases}$$

We focus on how AS approach the distribution of the total misreporting gap,

$$M_t = \sum_j M_{jt},$$

where $M_{nfp,t}^{NIPA} = M_{nfp,t}$. This implies that the full gap between total NIPA total income and tax income is:

$$G_t = M_t + D_{nfp,t}^{NIPA} + R_{nfp,t}.$$

The following mainly deals with the estimation of M_t, while $D_{nfp,t}^{NIPA}$ and $R_{nfp,t}$ are discussed at the end of this section.

Given this estimate of the overall misreporting gap, AS decompose this gap into several

[30] Throughout this section, all values directly reported by NIPA are indicated with the superscript "NIPA;" other variables are calculated/estimated by AS.

[31] Recall that DCE is based on the premise that the probability of observing a given dollar of income is less than one, but that the counterfactual where the probability is equal to one can be approximated using the variation in auditor effectiveness. If there are some group of auditors that are 100 percent effective in finding misreported income, and those auditors are assigned to tax units in some way uncorrelated with the probability of misreporting, then it is possible to construct the counterfactual where all misreported income is found. There are a number of critiques of this model - namely that auditors are not 100 percent effective and are not randomly assigned - that argue that the DCE methodology is either an under or over estimate of the misreporting gap.

components:

$$M_t = (M_t^D + M_t^U + M_t^{NF}) * S_t,$$

where $M_t^{NF} = 0.05 * M_t$ is the portion (five percent) of the gap allocated to non-filers, M_t^D is the portion of misreporting that can be detected via random audit data, M_t^U is the share that is undetected but inferred via DCE, and S is a scaling factor to preserve the equality between the left and right hand sides of the equation. The distributional implications of the misreporting gap is derived from estimates and assumptions about how these difference components are distributed.

To estimate both the size and distribution of detected misreporting (M_t^D), AS employed the methodology proposed by (Auten and Langetieg, 2023, henceforth AL) that uses information from random audit studies – theTCMP through 1988 and the NRP from 2001 through 2013 – to produce estimates of detected misreporting by year and reported income. AL tabulate the ratio of corrected income to reported income by year, rank of reported income, and - importantly - ratio groups, which are bins of the ratio of corrected to reported income. For each year and reported income percentile bin, AL report the share of tax units in each ratio group and summary statistics of the tax-unit-level ratio within that group. For example, one of the ratio groups contains observations with a ratio of corrected income to reported income of between 2 and 4. In 1988, AL estimate that 0.52 percent of tax returns between the 40th and 60th percentile of reported income were in this ratio group. For these returns in this ratio group, the average ratio is 2.56, and the "standard error for ratio" is 0.027 – whether this refers to the standard error of their estimate of the within-cell mean ratio or the estimated within-cell standard deviation of the ratio is unclear; the next step requires the latter but we cannot rule out an error here. The AL estimates contain many cell-level estimates of this kind. These estimates are the parameters of the micro-simulation we turn to next.

The steps of the micro-simulation used to distribute M_{jt}^D begins with tax-return level data, and does the following:

1. For each year, rank observations by reported income, and classify observations by their binned income percentile (p), where the bins match those reported by AL.
2. Randomly allocate observations within income percentiles to ratio groups (r) in proportion to the share of observations in that ratio group estimated by AL in the most recent wave of random audit data. Now each observation has been assigned an income percentile and ratio group.
3. Within each ratio group by income percentile by year cell, each observation is assigned a multiplier based on a random draw from that cell's ratio distribution, which is derived from the cell-level parameter estimates in AL.[32] ($m_{it}^D = m^D(p, r, \epsilon)$)
4. Each tax unit's detected-misreporting-corrected income is $y_{it} * m^D(p, r, \epsilon)$, so $M_{it}^D = y_{it} * (m^D(p, r, \epsilon) - 1)$

The total detected misreporting gap is the sum of the individual-level estimates of the dif-

[32] We are unable to locate a description of the specification of the assumed distribution in AS, AL, or appendices, apart from the fact that the distribution of the ratio in each cell is "bounded by max and min ratios, and the top group with a maximum of 125 percent of the cell mean."

ference between reported and corrected income:

$$M_t^D = \sum_i M_{it}^D.$$

After simulating values for detected misreporting, AS follow the process outlined in Auten and Splinter (2021) to simulate the size and distribution of DCE-identified undetected misreporting M_t^U. Their parametric specification, referred to as distributionally consistent gradient multipliers (DCGM), is proposed as a correction to the process used to distribute DCE adjustments in Johns and Slemrod (2010) and Mazur and Plumley (2007). This method entails multiplying detected under-reporting at the individual level M_{it}^D by a multiplier $m_{it}^{U,DCE}$ to allocate the additional under-reporting implied by DCE:

$$M_{it}^{U,DCE} = M_{it}^D * (m_{it}^{U,DCE} - 1) = y_{it} * (m^D(p, r, \epsilon) - 1) * (m^{U,DCE} - 1)$$

With the original approach of Johns and Slemrod (2010), the multiplier $m_{it}^{U,DCE}$ is essentially a constant calibrated from a different model in which auditor effects are actually estimated using 2001 NRP data.[33] With DCGM, AS allow the multiplier to vary with the ratio class (r) assigned above: $m^{U,DCE} = m^U(r)$.[34] They claim that the DCGM approach results in a more realistic distribution, because returns with large detected under-reporting would be more likely to have been assigned more effective auditors (so returns with higher r should get a lower multiplier $m^U(r)$). We are unable to ascertain exactly how AS specify the DSGM parameters ($m^U(r)$), but we discuss the way they validate their specified multipliers below. Unlike $m^D(p, r, e)$, $m^U(r)$ does not vary at the taxpayer level within a ratio class.

Using these ratio-specific multipliers, AS produce simulated values of each tax unit's undetected misreporting following the expression above. This is summed over all tax units to obtain the overall undetected share of misreporting.

$$M_t^U = \sum_i M_{it}^U$$

The last piece of this process is to add in the 5 percent of the gap assumed to be held by non-filers and scale the resulting estimate of the misreporting gap to match the observed gap between the tax and NIPA data.

Now we turn to the two additional components of the full NIPA gap and adjusted fiscal income, $D_{nfp,t}^{NIPA}$ and $R_{nfp,t}$. The latter is simply a residual, to make sure that the equation $G_{nfp,t} = M_{nfp,t}^{NIPA} + D_{nfp,t}^{NIPA} + R_{nfp,t}$ holds. The former is designed to account for the difference between economic depreciation and the tax deduction for depreciation. As discussed in the introduction, national accounts include an annual capital consumption adjustment for proprietor income to account for this difference; AS allocate the reported annual adjustment using two methodologies, depending on the period in question. For years prior to 1980, adjust-

[33]The Johns and Slemrod (2010) multiplier method, which was also used in the first draft of Guyton et al. (2023*b*), scales detected under-reporting by one of four constants depending on the type of income and whether the taxpayer was self-employed. See Guyton et al. (2023*a*) for further discussion; these authors describe this approach as DCE2001 because the multipliers are calibrated based on 2001 NRP data.

[34]To limit the effect of NOL carryovers, no multiplier is applied to tax units with negative AGI.

ments are allocated proportional to reported proprietor income. After 1980, first, 85 percent of reported expensing of sole proprietorships (and not partnerships) is added back to tax returns, then the remainder of the NIPA capital consumption adjustment is allocated proportionally to reported depreciation deductions of sole proprietorships (and not partnerships). The use of tax return variables pertaining to sole proprietor income to do this allocation rather than using sole proprietor and partnership income combined likely generates the bias described at the end of the introduction, which we do not discuss further in this comment.

We note that AS' paper is now forthcoming at the *Journal of Political Economy*, and it is unclear the extent to which the credibility of this method was assessed during peer review. The models above are critical to understanding how AS arrive at their distribution of misreported income, but important details about their methodology that AS used are described neither in the main body of their paper nor in the appendix to the paper. Rather, the micro-simulation parameters and their empirical basis is contained only in Auten and Langetieg (2023), which was not available publicly when the paper was accepted, while the approach to DCE is contained in a comment by Auten and Splinter on an earlier draft of Guyton et al. (2023*a*), i.e. Auten and Splinter (2021), which lies further into the weeds than referees would ordinarily be expected to wade. Neither of these were themselves subject to peer review.