Self-Employment Income Reporting on Surveys

Christian Imboden*
Georgia State University

John Voorheis*
U.S. Census Bureau

Caroline Weber*
University of Kentucky

May 2022

Abstract

This paper examines the relationship between tax administrative and survey income reports for self-employed individuals, relative to wage-earners, by comparing tax administrative income records directly from the IRS or via the Social Security Administration with responses to the Current Population Survey Annual Social and Economic Supplement (CPS-ASEC) from 2000 - 2015. We show that self-employed individuals report 51 percent more to the CPS-ASEC than to the IRS, on average. We observe a well-known artifact in the IRS data – self-employed individuals exhibit substantial bunching at the first EITC kink. When the survey and administrative data closely match (about 12 percent of the time), we observe the same artifact in the survey data. However, this artifact is not visible in the remaining portion of survey data. This suggests that differential reporting opportunities and incentives explain some of the differences between administrative and survey reports. The probability of close matches between the survey and administrative data increases among the self-employed sample when indicators of survey accuracy are present, which suggests survey accuracy may also play a role.

JEL Codes: C83, H24, H26.

Keywords: Income Reporting, Survey Accuracy, Measurement Error, Tax Evasion, Tax Avoidance

*Imboden: cimboden@gsu.edu; Voorheis: john.l.voorheis@census.gov; Weber: caroline.weber@uky.edu.
 The authors would like to thank David Agrawal, Chris Bollinger, Iuliia Shybalkina, Glen Waddell, and Jim Ziliak for helpful comments. We appreciate comments and feedback from participants at seminars at University of Kentucky and the National Tax Association Annual Conference. David Shi has provided exemplary research assistance.

1 Introduction

There is a long literature on the differences between wage-earners' (or all households') reports of earnings to surveys and tax authorities (e.g. Bound and Krueger, 1991; Bollinger, 1998; Pischke, 1995; Bound et al., 2001; Bollinger et al., 2019). This literature highlights several mismeasurement issues in the survey data, such as the "common man" hypothesis[1] and earnings non-response. Some more recent work allows for measurement error in administrative data as well, including mismatch, and finds less evidence of the common-man hypothesis (Kapteyn and Ypma, 2007; Abowd and Stinson, 2013; Meijer et al., 2012; Bollinger et al., 2018). Overall, these differences, while meaningful, generate relative small gaps between these two sources, on average. In contrast, we might expect households who derive most of their earnings from self-employment (i.e. as a sole proprietor) to potentially have a different relation between their survey and administrative earnings reports.

In this paper, we provide the first empirical examination of whether self-employed taxpayers have substantially different discrepancies in earnings between survey and tax administrative data, relative to wage-earners, and provide evidence on the sources of these differences. We use CPS-ASEC data linked to two tax administrative sources – Social Security Administration Detailed Earnings Records (DER) and some of the information contained from the IRS 1040. We describe our data in detail in Section 2. Our earnings measure is the sum of wage + self employment income for all taxpayers. This choice allows us to abstract from miscategorization issues previously documented in the literature, such as reporting self-employed income as wage income (Abraham et al., 2021).

[1]The common-man hypothesis is a desire to report earnings closer to the mean population earnings. So, if a taxpayer's earnings are below the mean, they will report higher earnings, and if their earnings are above the mean, they will report lower earnings.

We describe our empirical methods in Section 3 and our results in Section 4. We show much greater divergence between survey and administrative earnings reports for self-employed taxpayers, relative to wage-earners. Specifically, we find that self-employed individuals report 51 percent more to the CPS-ASEC than to the IRS, on average. And this difference is 44 percent larger than the same difference for wage-earners.

We examine several possible hypothesized explanations for the larger difference between survey and administrative data for self-employed households. First, self-employed and wage-earner households face substantially different incentives to engage in behavior to evade taxation because wage income faces informational reporting and withholding, whereas self-employment income does not.[2] IRS random audits suggest only 37 percent of income is reported for types of income, such as sole proprietorships, that are subject to little or no informational reporting. In contrast, 99 percent of wage-earnings are reported to the IRS (Slemrod and Bakija, 2008, p. 257). There are also additional tax avoidance opportunities available to self-employed households. To the extent that evasion or avoidance is not incorporated in survey reports, this will lead to a larger divergence of survey and administrative income for self-employed individuals. We identify a well-known artifact in administrative data to examine this hypothesis: bunching at the first EITC kink, which is driven in large part by evasion decisions (Chetty et al., 2013).[3] There are several key facts that emerge from this analysis. First, there is much less bunching at the first EITC kink in the survey data, and this is a mixture of two distributions; 12 percent have administrative and survey data that closely match and in these cases the bunching is approximately the same in the administra-

[2]Our discussion of tax evasion and avoidance in this paper is done exclusively to better understand what surveys measure when they ask questions about earnings. This paper cannot identify any individual case of tax avoidance or evasion.

[3]The first EITC kink occurs where the phase-in region ends and the plateau region begins in the EITC.

tive and survey data. For the remaining portion of the data, there is no observable bunching at the first EITC kink in the survey data. Second, relative to the administrative distribution, there is additional survey income mass both above and below the first EITC kink; this fact is consistent with evasion or avoidance being incorporated into administrative data, but not survey data. Third, for taxpayers that are married or have children (which moves the EITC kink rightward), the gap between administrative and survey records decreases.

Another hypothesis concerns the relative degree of survey accuracy for self-employed taxpayers relative to wage-earners: survey records might be less accurate for the self-employed who experience more income volatility and may be less certain of their income when asked. We find evidence that supports this hypothesis as well by examining how the likelihood of closely matched administrative and survey income reports and the percent gap between the two changes with measures of survey accuracy, such as whether survey income is rounded to the nearest $1,000 or whether the taxpayer files their IRS Form 1040 prior to March (the month of the CPS-ASEC). A plausible alternative interpretation of this evidence is that, as survey accuracy increases, the decision to avoid or evade taxation is more accurately reflected on the survey.

Our work contributes to several literatures. First, we contribute to the literature on the discrepancies between survey and administrative data that we discussed in our opening paragraph (Bee et al., 2021; Bee and Rothbaum, 2019). Our work examines the types of measurement error issues that are specific to self-employed households, an important subpopulation in the data, and benchmarks their measurement error issues to those of wage-earners. Second, we contribute to the literature on self-employed households (e.g. Bruce, 2000; Carroll et al., 2000; Gentry and Hubbard, 2000; Gale and Brown, 2013), their evasion

decisions across the income distribution (DeBacker et al., 2020) (excluding those at the very top of the income distribution), and whether evasion and avoidance is reported in surveys. We confirm the evidence in Hurst et al. (2014) that there is some tax evasion or avoidance in survey reports. We then make an additional contribution – there appears to be less evasion or avoidance in survey reports than in administrative records. Finally, we contribute to the literature on tax knowledge and tax rate perceptions (Gideon, 2017, e.g.): even for individuals that will bunch precisely at the first kink, they either don't know (partly because some haven't yet made the choice for those who haven't yet filed their tax return) or choose not to share with survey authorities these tax-related decisions.

2 Data

This study links data from one survey and two tax administrative sources covering tax years 2000 through 2015 (survey years 2001 through 2016). Our survey data come from the Current Population Survey Annual Social Economic Survey (CPS-ASEC) and our tax administrative data come from Social Security Administration Detailed Earnings Records (DER) and some items from the IRS Form 1040 tax returns. We describe the details of each component data source and the linkage procedure used to link records across the three data sets below.

The monthly CPS survey primarily collects employment information from about 60,000 American households, and is the source of the Bureau of Labor Statistics' published unemployment rates. One individual at a given address responds on behalf of all individuals living at that address. The address participates in the CPS for four months. Then, assuming the household has not changed addresses, they will participate again in the same four months

the following year. The ASEC is a supplement run each March that collects more detailed household income information and is the source of the Census Bureau's published poverty rates. Households may thus be surveyed in the ASEC for two consecutive years before being rotated out of the survey; this is called the overlap sample.

The ASEC collects individual information from most income categories, including income amounts that should also be recorded on Form 1040 (such as wages, self-employment earnings, interest, and dividends), as well as income amounts that are not recorded on Form 1040 (such as child support, SSI income and cash welfare assistance). Relative to the public use version of the CPS, the internal use version used here has higher income top codes. The CPS ASEC also collects demographic information such as household size, marital status, educational history, and occupation.

Our first tax administrative data source are the Detailed Earnings Records (DER), which are collected by the Social Security Administration from data they receive from the Internal Revenue Service. These detailed records contain all wage and self-employed earnings subject to Social Security and Medicare taxes (this is all positive non-farm and farm self-employment income from Form 1040 Schedules C & F income) for individuals that are in the Social Security system.[4] The Census Bureau receives an extract of the DER for most individuals who have responded to the CPS-ASEC since 1973.

Our second, supplementary, administrative data source provides additional information directly from the IRS Form 1040.[5] Detailed income information for all taxable categories and some non-taxable categories of income are reported on Form 1040. The U.S. Census

[4]Virtually all tax units pay into the Social Security system, but some are exempt due to religious objections or waivers.

[5]In this paper, Form 1040 refers to both the standard Form 1040 as well as variants such as the Form 1040A and Form 1040EZ.

Bureau receives some of the Form 1040 fields: wages and salaries, dividends received, both taxable and non-taxable interest income, social security income, rental and royalty income, total money income,[6], and adjusted gross income. We also receive information on filing status, number of dependents, and indicators (but not the amounts) for whether there was schedule C income (sole proprietorships), schedule D income (capital gains/losses), schedule E income (partnerships, LLCs, S corporations, rents and royalties) and Schedule F income (farms).

CPS-ASEC records are linked to DER and IRS 1040 records using the US Census Bureau's data linkage infrastructure. This infrastructure allows for the linking of individuals across survey and administrative records using anonymous identifiers called Protected Identity Keys (PIKs). PIKs are assigned to individuals in administrative records, surveys, or census microdata files using the Person Identification Validation System (PVS). PVS is a probabilistic matching algorithm which uses personally identifiable information (PII) to link individuals to a reference file. The reference files used by PVS is a modified version of the SSA's Numerical Identification File called the Census Numident. The Census Numident is the universe of individuals who have received Social Security Numbers (SSN), and contains PII including the SSN itself, as well as age, date of birth, sex, race and address.[7] PIKs are invariant across time and map one-to-one with SSNs. Once PIKs have been assigned to a file, it is possible to match with any other administrative records or survey records which have been assigned PIKs. We match all CPS ASEC respondents from survey years 2001 through 2016 with IRS 1040 and SSA DER data for the previous tax year (i.e. matching

[6]Total money income is our most complete measure of income that both contains self-employment earnings and is directly comparable between IRS (or SSA) and CPS sources.

[7]There are multiple vintages of the Numident reference file, each of which has the best PII information for a given individual.

the 2010 CPS ASEC to tax year 2009 IRS 1040s).

We drop any person records for whom we are not able to obtain a PIK and match across all three data sources (CPS, DER, and IRS). We aggregate the CPS and DER data to the tax-unit level when we match it with the IRS data.[8] Moreover, we drop any data points where the values in the CPS-ASEC were missing, imputed, or truncated.[9] We also drop any data points where zero wage + self-employment income was reported either to the DER or to the CPS. This last restriction is because our paper is focused exclusively on the intensive margin; that is, for tax-units that report positive income to both the CPS and DER, by how much more do these amounts differ for self-employed households relative to wage-earners? Because of these restrictions, our sample is not representative of the U.S. population as a whole, nor is it representative of the U.S. taxpaying population.

To identify self-employed and wage-earner taxpayers, we define taxpayers as "self-employed" when seventy-five percent or more of their administrative wages plus self-employment income comes from self-employment sources. All other individuals are "wage-earners." We aggregate up to the tax-unit or taxpayer level. This level of aggregation is natural for our EITC-related analysis below and our findings are not sensitive to this decision. Within taxpayers, we determine the primary earner and create additional demographic variables that describe the primary earner, including their age, gender, hours worked, education level (less than high school diploma completed, high school diploma completed with no college, some college completed with no degree, and bachelors degree completed), 2-digit reported occupation,

[8] Although we don't consider it in the analysis in this draft, we can also use only the CPS and DER data and conduct analysis at the individual taxpayer level (avoiding any aggregation).

[9] Specifically, we drop all taxpayer units with one or more individuals in the CPS that are either fully imputed or have imputed wage or self-employment income. When we analyze total money income, we drop all taxpayer units with one or more individuals in the CPS that are either fully imputed or have imputed income sources for any variable used in the construction of total money income in the CPS.

and their race/ethnicity (Black, Native American, Asian, Hispanic, White). We restrict the sample to those with a primary-earner aged 20 - 65 to focus on a working-age population.

We want to abstract away from basic miscategorization issues (i.e. reporting wage income as self-employment income and vice versa as documented in Abraham et al. (2021)), so our measure of earnings is wage + self-employment income (including farm income) in both the survey and administrative data.[10] Our measures of survey and administrative earnings are in real 2010 dollars using the PCE deflator (IRS's measure of inflation). We also consider a broader measure of income – total money income. This is a measure of income constructed from IRS records that is intended to match the sum of wages + self-employment income (including farm income) + interest + dividends + rental income + unemployment compensation + social security income + alimony + retirement income + other income in the CPS.

In addition to miscategorization of income in the survey data, it's also possible that there is miscategorization of income in the administrative data. Specifically, there is evidence that some individuals have self-employment income (as measured by 1099's), but do not report that income in a SE tax filing. Instead, they report the income as wage or other income (Collins et al., 2019). This form of miscategorization will not affect our total money income estimates of the gap between survey and administrative sources. If it's being categorized as other income, it will make these individuals to be more likely to be categorized as wage-earners, which will bias our estimates of the self-employed gaps, relative to wage earners, towards zero.

In the future, we intend to incorporate some additional data into our analysis. First, our

[10] We consider more elaborate miscategorization concerns in Section 4.

measures of EITC dependents and marital status will get more precise (by using a measure of EITC dependent eligibility instead of dependent for tax purposes and marital status based on IRS filing instead of CPS reports). We will also be able to observe the PVS codes that provide an indication of match quality to further examine the mismatch concerns we discuss below. And finally, we will incorporate a full history of DER records for all individuals to be able to more accurately characterize their current-year earnings relative to their average earnings to examine one form of survey accuracy that we begin to analyze below by looking at the overlap sample taxpayers' good, relative to bad, years.

3 Empirical Methods

Our analysis consists of two types of kernel densities (and means from these densities) as well as some regression analysis. In this section, we describe each. At many key junctures in our empirical analysis, we will compare the self-employed sample to the wage-earner sample. This will net out all forms of measurement error that are consistent across these two types of earners and allow us to follow the more recent literature that allows for measurement error in both survey and administrative sources (Kapteyn and Ypma, 2007; Abowd and Stinson, 2013; Meijer et al., 2012; Bollinger et al., 2018). Of course, it may be that certain measurement error issues are magnified among the self-employed sample; we will consider this as a possible explanation for our findings in our regression analysis.

We have two main forms of kernel densities presented in this paper. All kernel densities use a Gaussian kernel because of U.S. Census Bureau disclosure rules. The first main form of kernel density plots survey and administrative data separately for self-employed and wage-

earners (or occasionally for two categories of self-employed or wage-earners). The initial figures plot income in real 2010 dollars. Eventually, as noted in the text and figures, we switch to renormalized income. Renormalized income is real 2010 income that has been adjusted so that the first EITC kink occurs at the same point for all taxpayers regardless of marital status or the number of dependents. To do this, we choose one child unmarried households as our base unit of analysis.[11] Then, we multiply all households by the ratio of the first EITC kink point for one child unmarried households relative to the first kink point for the group to which that taxpaying unit belongs.

For some of our kernel densities, we reweight the wage-earner survey and administrative data to be more like the self-employed sample using inverse probability weighting. The first stage used to produce the weights regresses the following variables on a self-employment indicator: female, indicators for hours worked last week (under 10 hours, 60 hours and over, and indicators for every 5 hour increment in between, left out category is 35 to 40 hours), indicators for the number of dependents (0, 1, 2 or more), age and age squared, education indicators (less than high school, high school, some college, left out category is bachelor's degree), an indicator for being married, race and ethnicity indicators (Black, Asian, Native American, and Hispanic), and two-digit industry indicators.

Our second main form of kernel density estimate exploits the fact that we have linked survey and tax administrative data – we take the log difference of each taxpayer's administrative income minus their survey income (so a negative gap implies that their administrative income is lower than their survey income) and plot a kernel density of this difference for the

[11] In real 2010 dollars, the first EITC kink never moves for one child unmarried households during the period of our analysis. It is always $8,970.

self-employed and wage-earners.

Our regression equation is as follows:

$$y_{it} = \beta_0 + \beta_1 \cdot X_{it} + u_{it}, \tag{1}$$

where y_{it} is the outcome of interest for taxpayer i in year t (taxpayers in the overlap sample are observed in the data set twice). The outcome of interest is either an indicator for whether the taxpayer has closely matched (within \$2,500) survey and administrative income or the log difference of a taxpayer's administrative income minus their survey income. X_{it} are covariates of interest (demographics and survey accuracy), and u_{it} is a residual. When we benchmark our results relative to wage-earners, we include a self-employed indicator and interactions between this indicator and our covariates. To keep the tables concise, we report only the coefficients the interaction term coefficients. All standard errors are clustered by taxpayer.

4 Results

4.1 Kernel Density Analysis

We begin our analysis in the top panel of Figure 1 by plotting positive CPS (survey) and DER (tax administrative) wage + self-employment income separately by wage-earner and self-employed status for all taxpayers whose survey and administrative income is less than \$100,000.[12] The gap between self-employed survey and administrative data is substantial –

[12]We require both survey and administrative income to be below \$100,000 so that the same individuals appear in both figures. This restriction does not meaningfully affect the figure. The \$100,000 cut is imposed

survey income is an average of $13,000 (37 percent) higher than the administrative income.

In the bottom panel of Figure 1, we take the log difference of each taxpayer's administrative income minus their survey income and plot a kernel density of this difference for the self-employed (black line) and wage-earners (gray dashed line). This figure includes the same taxpayers as were included in the top panel. For wage-earners, there is a large spike at zero and the average gap is relatively small (-6.8 percent); in contrast, the peak of the density is negative for the self-employed and there is a long, fat left tail (the mean gap is -51.1 percent).

Figure 2 repeats this figure for total money, rather than wage + self-employment, income. This broader income definition is valuable to the degree that we are concerned about miscategorization of income. It's been previously documented in the literature that the self-employed categorize their self-employment income as wage income (Abraham et al., 2021), and it seems plausible that this could extend to other forms of income as well. For example, perhaps a taxpayer collects unemployment insurance or dividends, and reports those as wage or self-employment income on the CPS. These figures tell a similar story to the previous figure, though the gaps are slightly smaller, suggesting there is some merit to this concern. Alternatively it's possible that some of these additional categories are not as well-measured in the survey as in the administrative data and this accounts for the smaller gaps. Because the first EITC kink is defined in wage + self-employment income, we return to this narrower definition for the remaining analysis, but we'll return to this miscategorization issue in our regression analysis in Section 4.2.

---

because the density continues to thin out above $100,000 and so it becomes less interesting visually to include in this figure.

Figures 1 and 2 highlight much larger discrepancies between survey and administrative sources for the self-employed relative wage-earners. This result has important implications for any research studying self-employed earnings or poverty. The rest of the paper provides additional descriptive information and explores potential mechanisms behind this gap. The peak in administrative income in the region of the first EITC kink for the self-employed sample leads us to further exploration of that phenomenon as a potential mechanism.

The six panels in Figure 3 provide further descriptive details about the gap between survey and administrative data for the self-employed documented in Figure 1. Specifically, we provide three income splits of the bottom panel of Figure 1. We split by above/below \$50,000 in administrative income, above/below \$50,000 in survey income, and inside/outside the EITC range (based on administrative data). This leads to the following conclusions: almost all the detectable gap between administrative and survey data occurs below \$50,000 in administrative earnings; however it's much more equal across the survey earnings split. To the extent that administrative data incorporate more evasion and avoidance than survey data (which we'll explore in more detail below), evasion differences across income is one explanation for these patterns (DeBacker et al., 2020). The gap between survey and administrative data is larger for those within EITC range relative to outside of it, which is plausible if the first EITC kink is a focal point for evasion and avoidance activity among the self-employed and these activities are less likely to be incorporated into survey data. We explore this, as well as other plausible explanations, in more detail in the analysis that follows.

We now turn to documenting the behavior at the first EITC kink in more detail. As with most studies that examine potential tax avoidance or evasion, we cannot directly observe tax avoidance or evasion in the data; instead we are looking for statistical "Evidence of the

Invisible" (Slemrod and Weber, 2012). We know from existing literature (Saez, 2010; Chetty et al., 2013) that there is substantial bunching at the first EITC kink in tax administrative data for self-employed households (but much less, if any, for wage-earning households). Based on audit data, this is primarily due to tax evasion (Chetty et al., 2013). This is certainly not the only form of evasion or avoidance that occurs in administrative data; however, the salience of the first EITC kink makes it the easiest to utilize in this paper to examine whether, at least to some degree, administrative and survey reports diverge more among self-employed households because the self-employed engage in more tax evasion or avoidance relative to wage-earners and the modal choice is to not disclose evasion or avoidance adjustments when reporting income to survey authorities.

There are four panels in Figure 4. In the top left panel, we have renormalized taxpayers' earnings so that the first EITC kink of all taxpayers aligns. The remaining panels plot three unmarried samples: no children, one child, and two children households (which all share the same EITC first kink point by construction). The peak moves to the first EITC kink in each of the samples (the peak is slightly more diffuse for the no children sample, which isn't surprising given that this kink occurs at only $5,890 in real 2010 dollars). Survey income for these three samples stays more stable (and the mean is actually higher for those without children). As the first kink shifts right in these three panels, the average gap between survey and administrative reports declines.[13] This provides strong supporting evidence that one potential mechanism for the gap between survey and administrative data in self-employed households is tax evasion or avoidance. We more closely examine the moving kink by depen-

[13]The mean gap for wage-earners stays relatively constant across these groups: it is $1,000 for those without children and $2,000 for both groups with children.

dent and marital status in our regression analysis below. Another notable fact from these figures is that, in the samples with children, not only is there a substantial amount of missing mass above the first EITC kink in the administrative relative to survey data (which persists until well above $50,000), there is also missing mass below the first EITC kink. This is consistent with a story in which survey data reflect something closer to income before avoidance and evasion and administrative data include avoidance and evasion. This is because it is in taxpayers' best interest to overreport income if they would otherwise be below the first EITC kink, and underreport if they would otherwise be above the first EITC kink. For the rest of the analysis in this section, our baseline figure will be the renormalized combined figure in the top left panel.

Wage-earners and the self-employed may be different for numerous reasons. In Figure 5 we use the reweighting procedure described in Section 3 to reweight the wage-earner sample to look more like the self-employed sample based on observable covariates. We find that this does decrease the difference between wage-earner and self-employed income in the survey sample, but doesn't significantly alter the contrasting densities in the administrative data.

Next, we examine the frequency of "closely matched" survey and administrative data (within $2,500) for self-employed households in Figure 6 and Table 2. The top panel of Figure 6 separately plots the kernel density of survey and administrative records in cases in which these two sources closely match and when they don't closely match. The bottom panel repeats this plot for wage-earners for reference. For the self-employed, 12.2 percent of taxpayers are closely matched, relative to 24 percent for wage-earners. In the closely matched self-employed distribution, *both* the survey and administrative data have a spike at the first EITC kink, suggesting that, in these cases, any avoidance and evasion behavior

reported to administrative authorities is also precisely reported to survey authorities. If we take the administrative data as the relevant benchmark for wage-earners, as was done in the early literature on this topic (e.g. Bound and Krueger, 1991; Bollinger, 1998; Pischke, 1995; Bound et al., 2001), we can clearly see evidence of the "common-man" hypothesis in the wage-earner figure for the non-matched sample – at the bottom of the income distribution, survey earnings exceed administrative earnings, and then this switches around $40,000. But such behavior is not observable in the self-employed sample; instead, for the unmatched sample, the administrative density peaks at the first EITC kink and after that peak declines, the mass of the administrative data remains forever below the mass in the survey data.

4.2 Regression Analysis

In this section, we turn to regression analysis to more closely analyze possible explanations for the large gaps between survey and administrative data for the self-employed using equation (1). Before we begin our regression analysis, we provide summary statistics on our dependent variables and covariates in Table 1. For our regression analysis, we examine the region of administrative income within $5,000 of the first EITC kink. This focuses our analysis around the first kink and holds administrative income relatively constant. This minimizes the degree to which income levels may confound our relations of interest.[14]

The summary statistics we provide in Table 1 match our regression sample and provide information on our regression dependent variables and covariates. All summary statistics are displayed separately for self-employed and wage-earners. About one-fifth of the taxpayers in

[14]Our conclusions are quite similar if we examine the full distribution that we've considered in our kernel density figures.

this analysis are self-employed. Our first variable – percent self-employed – demonstrates the validity of our self-employed and wage-earner definitions. About 98% of earnings are self-employment earnings in the self-employed sample, and only about 4% of earnings are self-employment earnings in the wage-earner sample. This particular region of the administrative earnings distributions has slightly higher rates of administrative and survey earnings closely matching (within $2,500 of each other) relative to the full distribution – about 18% of the self-employed sample closely match and twice as many wage-earners closely match. This area is also an area with larger negative percent administrative - survey gaps: both are substantially negative, but the self-employed gap is about twice that of the wage-earner gap. Please see the table for means of all covariates we use in our analysis.

Table 2 examines what predicts having closely matched survey and administrative earnings amounts and Table 3 examines what predicts the size of the percent gap of administrative - survey income. Columns (1) - (2) examine self-employed taxpayers by themselves. Columns (3) and (4) examine the additional ability of these covariates to explain differences for the self-employed, relative to the wage-earner, sample by including wage-earners in the regression and reporting the interactions from a self-employed indicator and our covariates.

Our first set of covariates are demographic ones – an indicator for being female, married, having children, over 40 years old, and education indicators (high school degree, some college, bachelor's degree). It's well-documented in the wage-earner literature that the differences between survey and administrative sources are lower for women – we see that for the self-employed as well in Column (1) of Tables 2 and 3. Relative to wage-earners in Column (3), though, these effects are insignificant (or even slightly negative). The next two covariates: marital status and whether or not there are dependents in the household are demographics

that directly move the first EITC kink as we examined in Figure 4. As one either gets married or has dependents, the kink moves rightward; thus, as we discuss above, for a given amount of true income, the amount of avoidance or evasion needed is lower. Our results are consistent with this – relative to wage-earners, the match rate increases and the gap between the two income records declines when individuals are married or have dependents. The coefficients on age and education are broadly consistent with gaps increasing (in absolute magnitude) with proxies of permanent income (which is consistent with the survey income behavior documented in Figure 3 to the extent that survey income is a better proxy for permanent income than is administrative income).

The remaining covariates are related to the relative accuracy of the two income records. Those who report being full-time last week may not have been fully-employed for the whole year, but may provide answers to the CPS that are more consistent with their last week earnings (and this may be magnified for the self-employed). The regression estimates support this hypothesis.

Because the CPS-ASEC does not ask about most of Schedule E and Schedule F can be difficult to capture on a survey for numerous reasons, we include an indicator for whether the taxpayer had any schedule E or F income (only in Columns (1) and (2)). Match rates are higher and gaps are lower in absolute magnitude when there is no Schedule E or F income.

We might also worry about miscategorization of income as we have already considered in Figure 2. To examine this issue here, we include an indicator for administrative wage + self-employment income being within 10% of total money income (which includes these other forms of income, such as dividends and unemployment insurance). The earnings gaps between these two sources are substantially smaller in absolute magnitude when wage +

self-employment income is within 10 percent of total money income in Table 3, though it does not appear to be significantly magnified for the self-employed relative to wage-earners.

Mismatch is another concern that has been previously documented in the literature and could theoretically have a substantial effect on our analysis – perhaps self-employed households sometimes get matched to someone else (who often happens to not be self-employed since the self-employed are a relatively small fraction of the population). In many of these cases, it's plausible to expect that the self-employed taxpayer bunched at the first EITC kink in the administrative data and then their matched wage-earner counterpart did not bunch in the survey data (because they weren't bunching in either the survey or administrative data), and this created the discrepancy we have documented here. As one way of examining this hypothesis, we consider an indicator for whether the person believes they are self-employed in the survey as well (i.e. 75 percent of their income in the survey is identified as self-employment income). For these taxpayers, the likelihood of this mismatch issue being present is much lower. Self-employed taxpayers who also report being self-employed to the survey are more likely for their two income records to closely match and the average gaps between the two are smaller in absolute magnitude. There are two possible interpretations of these coefficients. The first is that individuals who are more likely to be an accurate match (as indicated by thinking they are self-employed in the survey and being self-employed in the administrative data), are more likely to have closely matched survey and administrative income sources. The second possible story is simply one of accuracy – perhaps taxpayers who are careful enough to accurately report their self-employment income as self-employment income in the survey are more accurate in their reporting of total amounts as well.

We find that filing an IRS Form 1040 before the survey increases the likelihood that

survey and administrative incomes closely match and decreases the absolute magnitude of the gap between the two sources (and the latter is magnified for the self-employed relative to wage-earners).[15] There are two possible interpretations. First, reporting to tax authorities first forces individuals to make their avoidance and evasion decisions (which they may not have previously considered) and then these are then also more likely to be reported to survey authorities. The second interpretation is that if a taxpayer has not reported their income to tax authorities yet, they may have substantial uncertainty about their income (and this is magnified for self-employed households because they have more income uncertainty, in general). So, taxpayers might round up their income in the survey and this is part of what pushes survey income away from the first EITC kink.

We do find evidence that the decision not to round income to the nearest $1,000 does increase the match rate by a substantial amount (but not relative to wage-earners) and does decrease the gap in absolute magnitude (even relative to wage-earners) – this lends some support to this alternative theory that self-employed taxpayers round up substantially when they round.[16] Though, it could also be that individuals who decide to be accurate incorporate their evasion and avoidance decisions, while those that do not choose the same degree of accuracy in their survey reporting are also more likely to leave out their avoidance and evasion when responding to the survey.

Returning to the "common-man hypothesis," we would expect that taxpayers would have the strongest incentives to round up substantially when their income is lower, both because they might be ashamed to report the low amount and also because they might

[15]We define "filing before the survey" as having their Form 1040 processed in the first 9 weeks of the year. All CPS-ASEC interviews take place in the month of March.

[16]Note, though, that it is not a mechanical effect. We can round the administrative data in the same was as the survey data is rounded for each individual and this does not have a large effect.

instinctively report their average, rather than their year-to-year, income. In Columns (2) and (4), we restrict our analysis to the overlap sample and add an indicator for whether it's the taxpayer's "good year" (i.e. their higher income year according to administrative data) in the overlap sample. This has little effect on the likelihood of a close match, but does decrease the absolute gap in earnings between the two sources. If we look at Figure 7, which splits self-employed taxpayers by their good year vs. bad year in the top panel and repeats the same for wage-earners in the bottom panel, it appears this result is being driven by a larger move in administrative than survey income between these two years. Our final regression coefficient suggests that there don't appear to be important differences depending on whether it's the first or second year responding to the survey.

In conclusion, we have strong evidence that survey accuracy as well as avoidance and evasion both potentially play important roles in the relative reports to survey and tax authorities.

5 Conclusion

We often study self-employed taxpayers and their outcomes in either survey or administrative data (e.g. Bruce, 2000; Carroll et al., 2000; Gentry and Hubbard, 2000; Gale and Brown, 2013). And yet we know relatively little about how these two sets of income reports compare and why, because there has not been a comprehensive intensive margin study of survey measurement among this sub-population. In this study, we abstract away from previously documented miscategorization issues (e.g. self-employed individuals declaring that they make wage income and vice versa), and compare the wage + self-employment income of the

self-employed relative to their wage-earning counterparts in linked survey and administrative data. Our analysis is focused on individuals that report positive amounts of income to both survey and administrative sources (and is thus not representative of the survey or administrative populations). We find dramatically larger gaps for the self-employed; they report 51 percent more to the survey than they report to the IRS, on average. The fact that the means of the survey and administrative data are so far apart for self-employed households indicates that the choice of income source will matter a lot for most analysis of self-employed taxpayers, for the determination of poverty rates, and so forth.

Identifying the mechanisms behind this larger gap among the self-employed is a particularly challenging problem as it's quite plausible that both the survey and administrative data sources are biased. Nevertheless, it's a very important question; it's an essential step toward a better understanding of survey measurement quality for the self-employed and we hope this increased understanding will aid in survey measurement improvements for this sub-population in the years to come. One possible mechanism is that some of the well-documented survey measurement issues for wage-earner households are exacerbated among the self-employed. We find some evidence that this is indeed the case. We consider several measures of survey accuracy; each significantly decreases the gap for the self-employed, and many significantly decrease the gap more for the self-employed than for wage-earners.

It's also the case that the administrative data, by construction, reflects the tax avoidance and evasion choices of all taxpayers. It's well known that, because of the structure of the tax code and lack of informational reporting and withholding, the self-employed engage more frequently in tax avoidance and evasion activities. To the extent that these choices are not also reflected in survey earnings reports, this provides another plausible mechanism for the

larger difference between these two sources among self-employed taxpayers. We find evidence of this as well. Specifically, the first EITC kink, a well-known focal point of tax evasion in the administrative tax data, is a key driver of this gap – the gap falls by more than half for taxpayers not in the EITC range. Moreover, as the number of children increases, the peak in administrative data moves rightward and the gap between administrative and survey income declines. This potentially suggests that a substantial portion of avoidance or evasion reported in the administrative data is not reflected in the survey data.

References

Abowd, John M. and Martha H. Stinson, "Estimating Measurement Error in Annual Job Earnings: A Comparison of Survey and Administrative Data," *The Review of Economics and Statistics*, 2013, *95* (5), 1451–1467.

Abraham, Katharine G., John C. Haltiwanger, Claire Hou, Kristin Sandusky, and James R. Spletzer, "Reconciling Survey and Administrative Measures of Self-Employment," *Journal of Labor Economics*, 2021, *39* (4), 825–860.

Bee, Adam and Jonathan Rothbaum, "The Administrative Income Statistics (AIS) Project: Research on the Use of Administrative Records to Improve Income and Resource Estimates," SEHSD Working Paper Number 2019-36 2019.

_ , Sandy Clark, Kevin Rinz, Jonathan Rothbaum, Chase Sawyer, John Voorheis, Thomas Hertz, Pat Langetieg, and Mark Payne, "Improving Income Measurement in the American Community Survey Using Expanded 6103(n) Income Data from the Internal Revenue Service," Mimeo 2021.

Bollinger, Chris R., Barry T. Hirsch, Charles M. Hokayem, and James P. Ziliak, "The Good, The Bad and The Ugly: Measurement Error, Non-response and Administrative Mismatch in the CPS," Working Paper 2018.

Bollinger, Christopher R., "Measurement Error in the Current Population Survey: A Nonparametric Look," *Journal of Labor Economics*, 1998, *16* (3), 576–594.

_ , Barry T. Hirsch, Charles M. Hokayem, and James P. Ziliak, "Trouble in the Tails? What We Know about Earnings Nonresponse 30 Years after Lillard, Smith, and Welch," *Journal of Political Economy*, 2019, *127* (5), 2143–2185.

Bound, John and Alan B. Krueger, "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?," *Journal of Labor Economics*, 1991, *9* (1), 1–24.

_ , Charles Brown, and Nancy Mathiowetz, "Chapter 59 - Measurement Error in Survey Data," in James J. Heckman and Edward Leamer, eds., *James J. Heckman and Edward Leamer, eds.*, Vol. 5 of *Handbook of Econometrics*, Elsevier, 2001, pp. 3705–3843.

Bruce, Donald, "Effects of the United States Tax System on Transitions into Self-Employment," *Labour Economics*, 2000, *7* (5), 545–574.

Carroll, Robert, Douglas Holtz-Eakin, Mark Rider, and Harvey S. Rosen, "Income Taxes and Entrepreneurs' Use of Labor," *Journal of Labor Economics*, 2000, *18* (2), 324–351.

Chetty, Raj, John N. Friedman, and Emmanuel Saez, "Using Differences in Knowledge across Neighborhoods to Uncover the Impacts of the EITC on Earnings," *American Economic Review*, December 2013, *103* (7), 2683–2721.

Collins, Brett, Andrew Garin, Emilie Jackson, Dmitri Koustas, and Mark Payne, "Is Gig Work Replacing Traditional Employment? Evidence from Two Decades of Tax Returns," *Working Paper*, 2019.

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

Gale, William and Samuel Brown, "Small Business, Innovation, and Tax Policy: A Review," *National Tax Journal*, 2013, *66* (4), 871–892.

Gentry, William M. and R. Glenn Hubbard, "Tax Policy and Entrepreneurial Entry," *American Economic Review*, May 2000, *90* (2), 283–287.

Gideon, Michael, "Do Individuals Perceive Income Tax Rates Correctly?," *Public Finance Review*, 2017, *45* (1), 97–117. PMID: 29238156.

Hurst, Erik, Geng Li, and Benjamin Pugsley, "Are Household Surveys Like Tax Forms? Evidence from Income Underreporting of the Self-Employed," *The Review of Economics and Statistics*, 03 2014, *96* (1), 19–33.

Kapteyn, Arie and Jelmer Y. Ypma, "Measurement Error and Misclassification: A Comparison of Survey and Administrative Data," *Journal of Labor Economics*, 2007, *25* (3), 513–551.

Meijer, Erik, Susann Rohwedder, and Tom Wansbeek, "Measurement Error in Earnings Data: Using a Mixture Model Approach to Combine Survey and Register Data," *Journal of Business and Economic Statistics*, 2012, *30*, 191–201.

Pischke, Jörn-Steffen, "Measurement Error and Earnings Dynamics: Some Estimates from the PSID Validation Study," *Journal of Business Economic Statistics*, 1995, *13* (3), 305–314.

Saez, Emmanuel, "Do Taxpayers Bunch at Kink Points?," *American Economic Journal: Economic Policy*, August 2010, *2* (3), 180–212.

Slemrod, Joel and Caroline Weber, "Evidence of the invisible: Toward a credibility revolution in the empirical analysis of tax evasion and the informal economy," *International Tax and Public Finance*, 2012, *19* (1), 25–53.

_ **and Jon Bakija**, *Taxing Ourselves*, 4th ed., Cambridge, Massachusetts: The MIT Press, 2008.

Tables

Table 1: Summary Statistics

	Self-Employed		Wage-Earners	
Variable	**Obs.**	**Mean**	**Obs.**	**Mean**
Percent Self-Employment Earnings	11,000	0.977	52,500	0.037
Closely Match	11,000	0.177	52,500	0.356
Percent DER - CPS Gap	11,000	−0.791	52,500	−0.447
Female	11,000	0.425	52,500	0.594
Dependents	11,000	0.608	52,500	0.647
Married	11,000	0.557	52,500	0.503
Age over 40	11,000	0.594	52,500	0.449
High-School Degree Only	11,000	0.364	52,500	0.373
Some College	11,000	0.266	52,500	0.303
Bachelor's Degree	11,000	0.177	52,500	0.149
Full-Time	11,000	0.584	52,500	0.480
No Schedule E or F Income	11,000	0.809	52,500	0.910
Abs(TMI - Wage + SE)<10%	11,000	0.664	52,500	0.578
Survey Self-Employed	11,000	0.410	52,500	0.032
IRS File Before CPS	11,000	0.317	52,500	0.633
Not $1,000 Round	11,000	0.172	52,500	0.277
Good Year	5,300	0.424	23,000	0.392
Second Year	5,300	0.499	23,000	0.488

These are summary statistics for our regression sample: within $5,000 of the EITC kink in administrative earnings and above $1,000 in administrative and survey earnings (so the not round covariate is well-defined).

Table 2: Closely-Matched Earnings Table

	(1)	(2)	(3)	(4)
Self-Employed			−0.265***	−0.282***
			(0.036)	(0.053)
Female	0.048***	0.069***	−0.046***	−0.041**
	(0.008)	(0.013)	(0.009)	(0.014)
Dependents	0.043***	0.057***	0.049***	0.069***
	(0.008)	(0.011)	(0.009)	(0.013)
Married	−0.040***	−0.047***	0.013	−0.007
	(0.008)	(0.012)	(0.009)	(0.014)
Age over 40	0.014	0.016	0.012	0.012
	(0.008)	(0.012)	(0.009)	(0.013)
High-School Degree Only	−0.021*	−0.02	0.004	0.009
	(0.011)	(0.016)	(0.012)	(0.019)
Some College	−0.028*	−0.048**	0.01	−0.005
	(0.012)	(0.017)	(0.013)	(0.019)
Bachelor's Degree	−0.027*	−0.044*	0.029*	0.024
	(0.012)	(0.018)	(0.014)	(0.021)
Full-Time	−0.085***	−0.080***	0.034***	0.046***
	(0.008)	(0.011)	(0.009)	(0.013)
No Schedule E or F Income	0.022*	0.028*		
	(0.009)	(0.013)		
Abs(TMI - Wage +SE)<10%	0.025**	0.01	−0.046***	−0.068***
	(0.008)	(0.012)	(0.009)	(0.013)
Survey Self-Employed	0.064***	0.064***		
	(0.007)	(0.011)		
IRS File before CPS	0.064***	0.074***	0.017	0.026
	(0.009)	(0.013)	(0.010)	(0.015)
Not $1,000 Round	0.117***	0.134***	−0.044***	−0.032
	(0.011)	(0.016)	(0.012)	(0.018)
Good Year		0.006		−0.002
		(0.010)		(0.011)
Second Year		−0.014		−0.003
		(0.009)		(0.011)
Observations	11,000	5,400	64,000	28,500
Restricted to Overlap Sample	No	Yes	No	Yes
Includes Wage-Earners	No	No	Yes	Yes

Standard errors are in parentheses. Standard errors are clustered by taxpayer. *10% significance level. **5% significance level. ***1% significance level. The dependent variable in this regression is an indicator for survey and administrative earnings being "closely matched" (within $2,500 of each other). Our regression includes taxpayers within $5,000 of the EITC kink in administrative earnings and above $1,000 in administrative and survey earnings (so the not round covariate is well-defined). Columns (1) and (2) are for self-employed individuals only; Columns (3) and (4) include wage-earners as well and report the interaction terms on a self-employed indicator and the covariate specified (so that the coefficients can be interpreted as the additional effect of the covariate for the self-employed, relative to wage-earner, sample. Columns (2) and (4) restrict to the overlap sample – the sample of taxpayers observed in two adjacent years in the CPS – and variables that are only defined for this group are added.

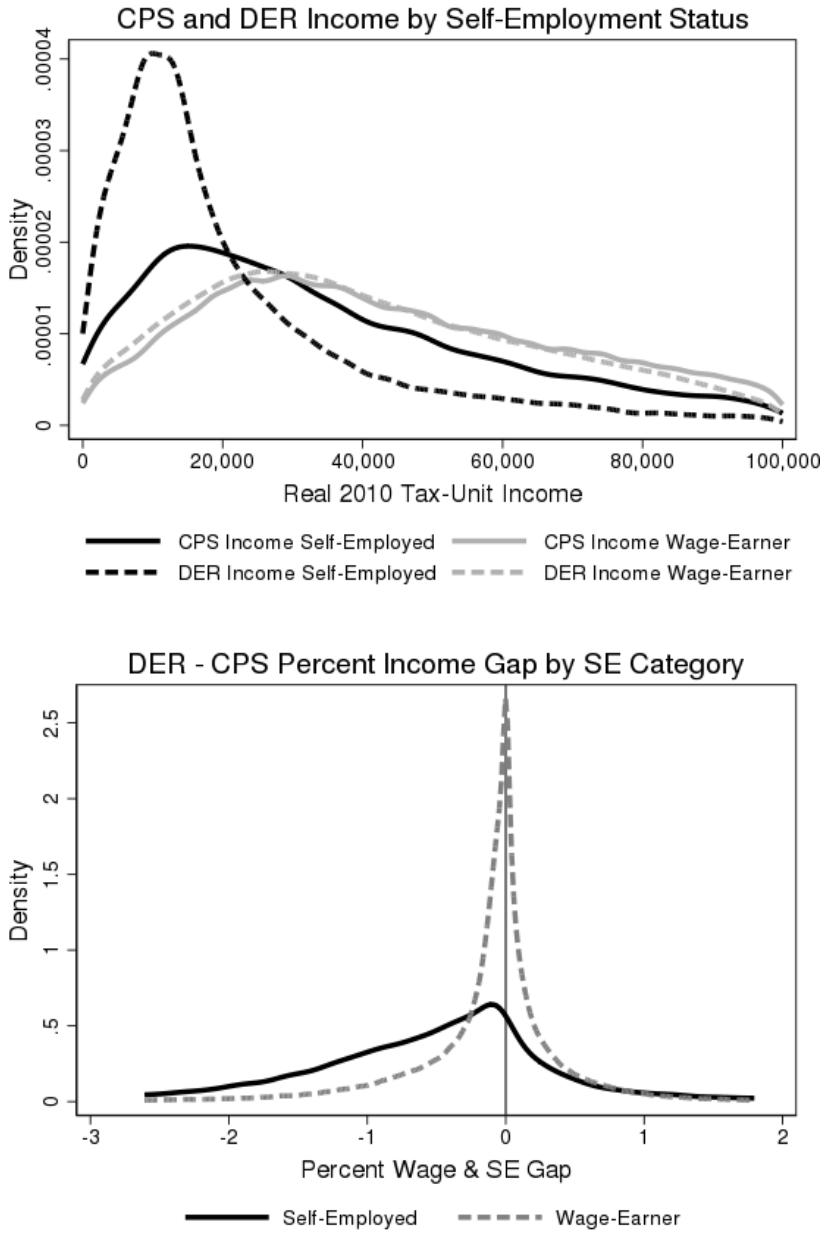
Table 3: Percent Administrative - Survey Earnings Table

	(1)	(2)	(3)	(4)
Self-Employed			−0.265***	−0.282***
			(0.036)	(0.053)
Female	0.268***	0.292***	−0.025	−0.04
	(0.018)	(0.027)	(0.020)	(0.029)
Dependents	0.421***	0.412***	0.246***	0.245***
	(0.018)	(0.026)	(0.020)	(0.029)
Married	0.040*	0.062*	−0.023	−0.014
	(0.018)	(0.026)	(0.020)	(0.029)
Age over 40	−0.092***	−0.088***	−0.048**	−0.047
	(0.017)	(0.025)	(0.018)	(0.027)
High-School Degree Only	−0.174***	−0.208***	−0.037	−0.071*
	(0.022)	(0.030)	(0.023)	(0.033)
Some College	−0.258***	−0.291***	−0.054*	−0.089*
	(0.024)	(0.033)	(0.026)	(0.036)
Bachelor's Degree	−0.440***	−0.470***	−0.024	−0.006
	(0.028)	(0.039)	(0.031)	(0.044)
Full-Time	−0.389***	−0.365***	−0.069***	−0.042
	(0.017)	(0.024)	(0.018)	(0.026)
No Schedule E or F Income	0.186***	0.173***		
	(0.025)	(0.036)		
Abs(TMI - Wage +SE)<10%	0.186***	0.199***	0.034	0.018
	(0.020)	(0.029)	(0.020)	(0.029)
Survey Self-Employed	0.155***	0.173***		
	(0.016)	(0.023)		
IRS File before CPS	0.217***	0.205***	0.062**	0.058*
	(0.018)	(0.025)	(0.020)	(0.028)
Not $1,000 Round	0.541***	0.502***	0.131***	0.113***
	(0.023)	(0.031)	(0.024)	(0.033)
Good Year		0.127***		0.125***
		(0.021)		(0.023)
Second Year		−0.022		−0.033
		(0.020)		(0.023)
Observations	11,000	5,400	64,000	28,500
Restricted to Overlap Sample	No	Yes	No	Yes
Includes Wage-Earners	No	No	Yes	Yes

Standard errors are in parentheses. Standard errors are clustered by taxpayer. *10% significance level. **5% significance level. ***1% significance level. The dependent variable in this regression is the percent gap in administrative - survey earnings. Our regression includes taxpayers within $5,000 of the EITC kink in administrative earnings and above $1,000 in administrative and survey earnings (so the not round covariate is well-defined). Columns (1) and (2) are for self-employed individuals only; Columns (3) and (4) include wage-earners as well and report the interaction terms on a self-employed indicator and the covariate specified (so that the coefficients can be interpreted as the additional effect of the covariate for the self-employed, relative to wage-earner, sample. Columns (2) and (4) restrict to the overlap sample – the sample of taxpayers observed in two adjacent years in the CPS – and variables that are only defined for this group are added.

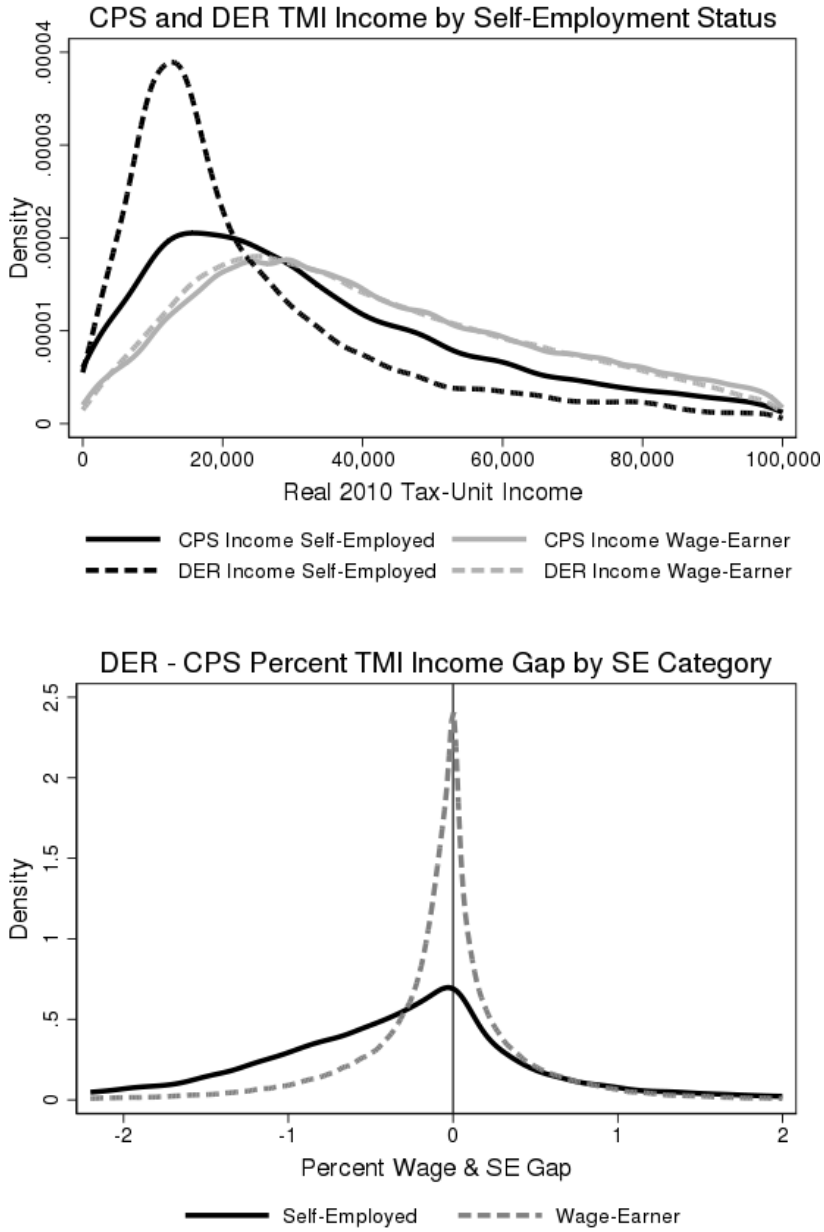
Figures

Figure 1: Kernel Density Plots by Self-Employment Status



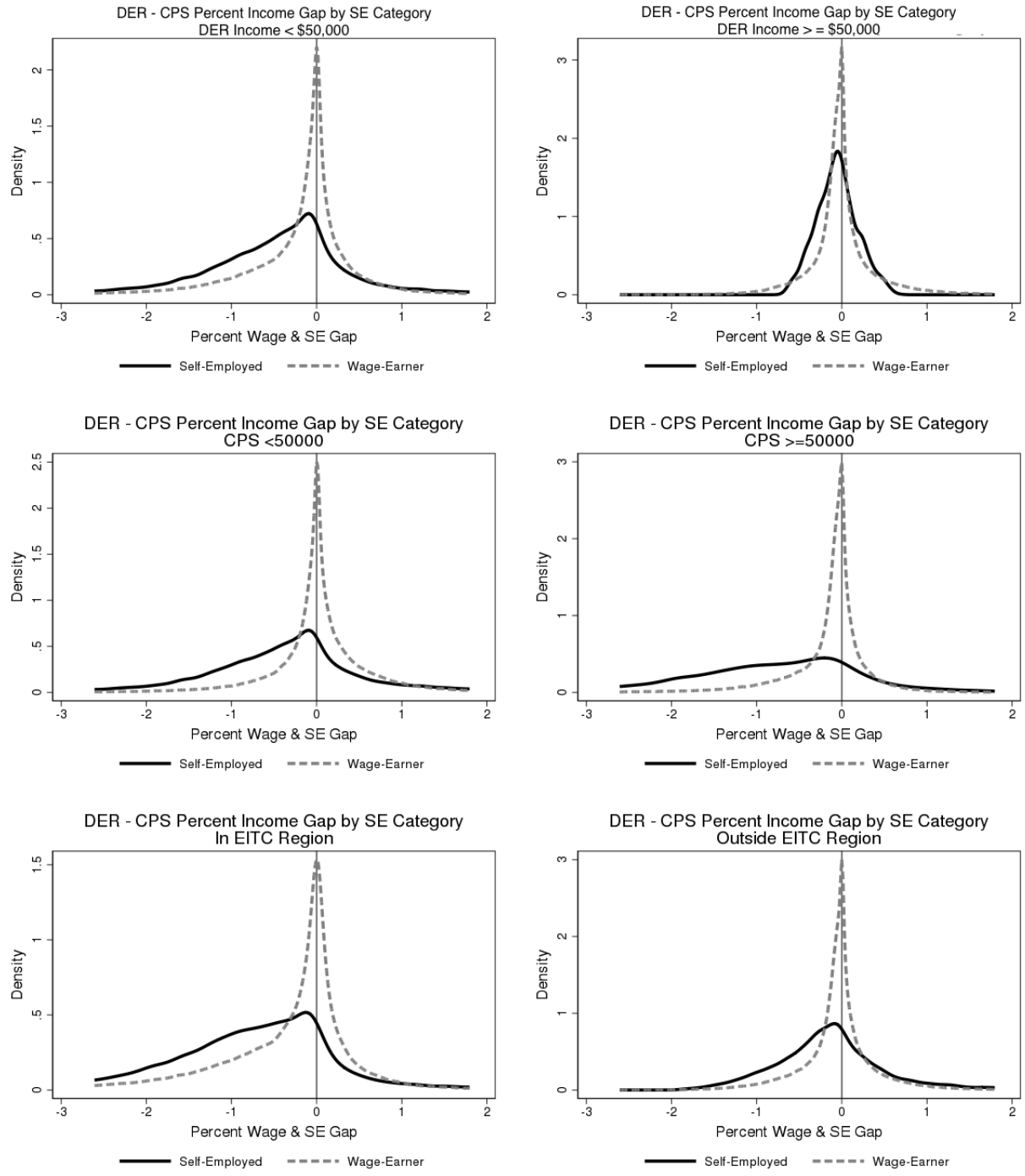
The top figure plots a Gaussian kernel of positive survey (CPS, solid lines) and administrative (DER, dashed lines) income for all taxpayers whose survey and administrative income is less than $100,000 separately by self-employed (black lines) and wage-earners (gray lines). We require both survey and administrative income to be below $100,000 so that the same individuals appear in both figures. This restriction does not meaningfully affect the figure. Average survey earnings for self-employed individuals is $35,000, relative to their administrative earnings of $22,000. In contrast, wage earnings are relatively close – $44,000 in the survey and $41,000 in the administrative data. The bottom panel subtracts the log of individual's administrative income from the log of their survey income (so a negative gap implies that their administrative income is lower than their survey income) and plots a kernel density of this difference for the self-employed (black line) and wage-earners (gray dashed line). The average gap for the self-employed is -51.1%; the average gap for wage-earners is -6.8%.

Figure 2: Total Money Income Kernel Density Plots by Self-Employment Status



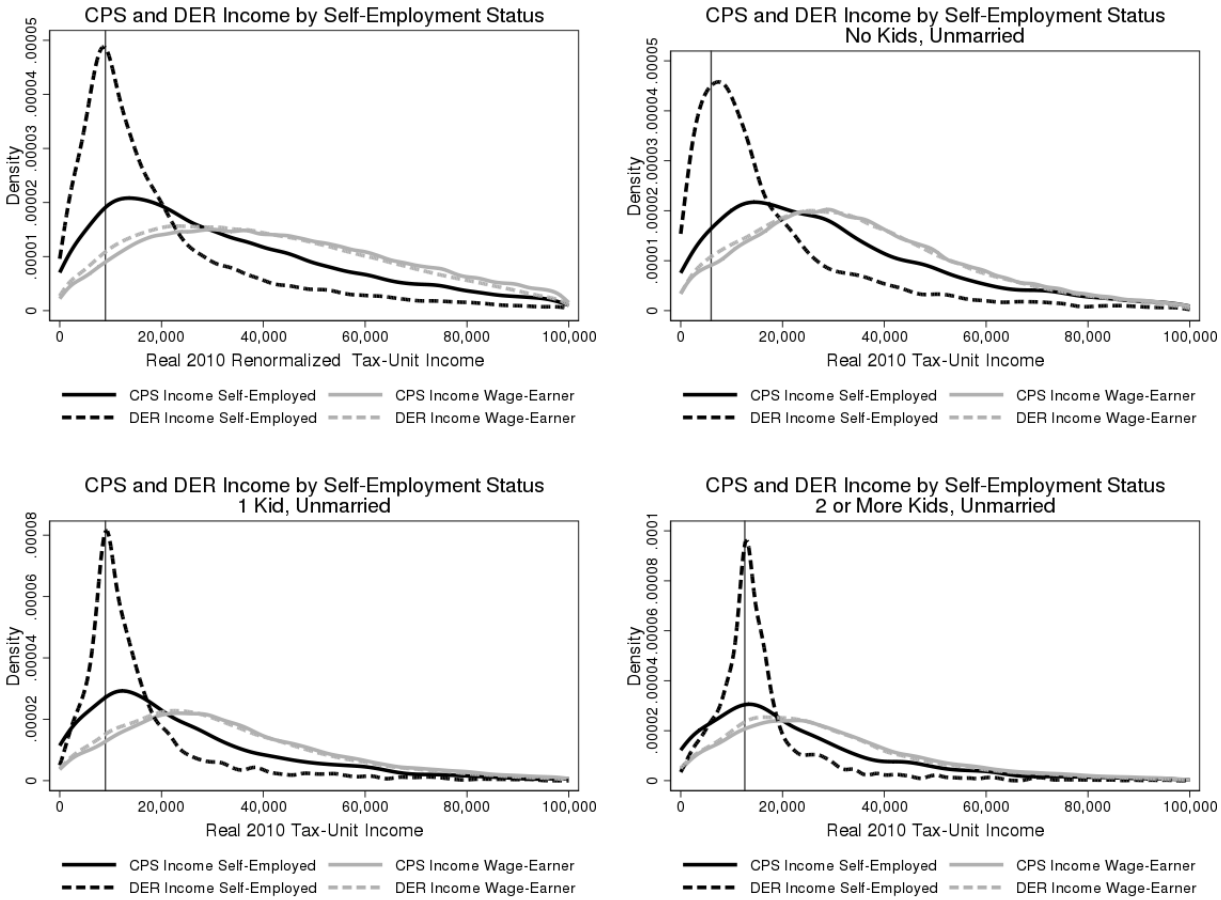
This figure repeats Figure 1 for total money income, instead of wage + self-employment income. Average survey earnings for self-employed individuals is $34,000, relative to their administrative earnings of $25,000. In contrast, wage earnings are relatively close – $42,000 in the survey and $41,000 in the administrative data. The bottom panel subtracts the log of individual's administrative income from the log of their survey income (so a negative gap implies that their administrative income is lower than their survey income) and plots a kernel density of this difference for the self-employed (black line) and wage-earners (gray dashed line). The average gap for the self-employed is -32.3%; the average gap for wage-earners is -0.031%.

Figure 3: Kernel Density Plots of Percent Income Gap by Self-Employment Status



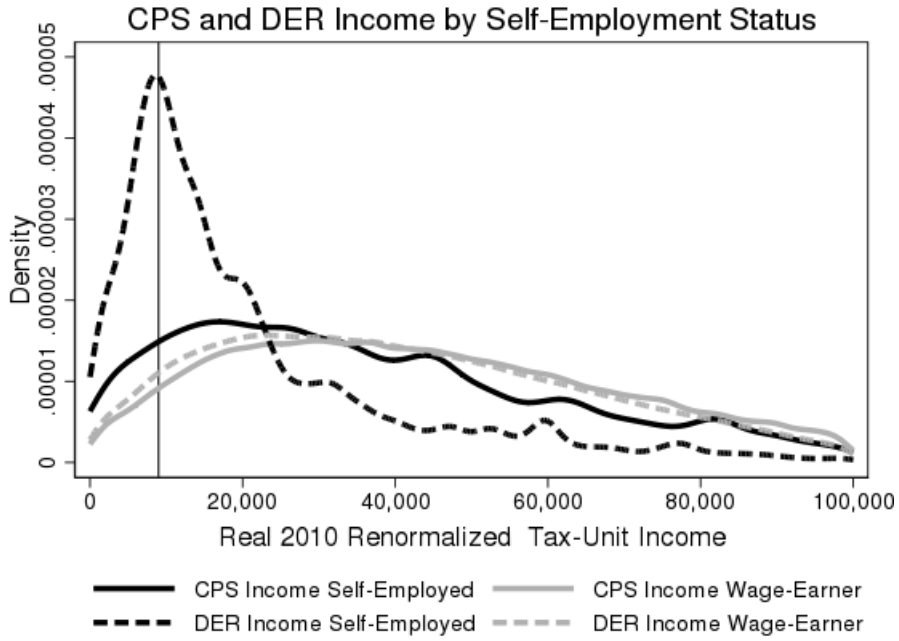
Each panel subtracts the log of individual's administrative income from the log of their survey income (so a negative gap implies that their administrative income is lower than their survey income) and plots a kernel density of this difference for the self-employed (black line) and wage-earners (gray dashed line). The top two panels split by having administrative income below (left) or above (right) $50,000. The mean difference in the self-employed minus wage-earner distributions for the left top panel is -0.40 (p-value=0.000) and the mean difference in the right top panel is -0.03 (p-value=0.000). The middle two panels repeat this split by survey income levels. The mean difference in the left middle panel is -0.40 (p-value=0.000) and the mean difference in the right middle panel is -0.56 (p-value=0.000). And the bottom left panel is for all individuals with administrative income in the EITC region and the bottom right panel is for all individuals outside the EITC region. The mean difference in the left bottom panel is -0.42 (p-value=0.000) and the mean difference in the right middle panel is -0.17 (p-value=0.000).

Figure 4: First EITC Kink Kernel Density Plots



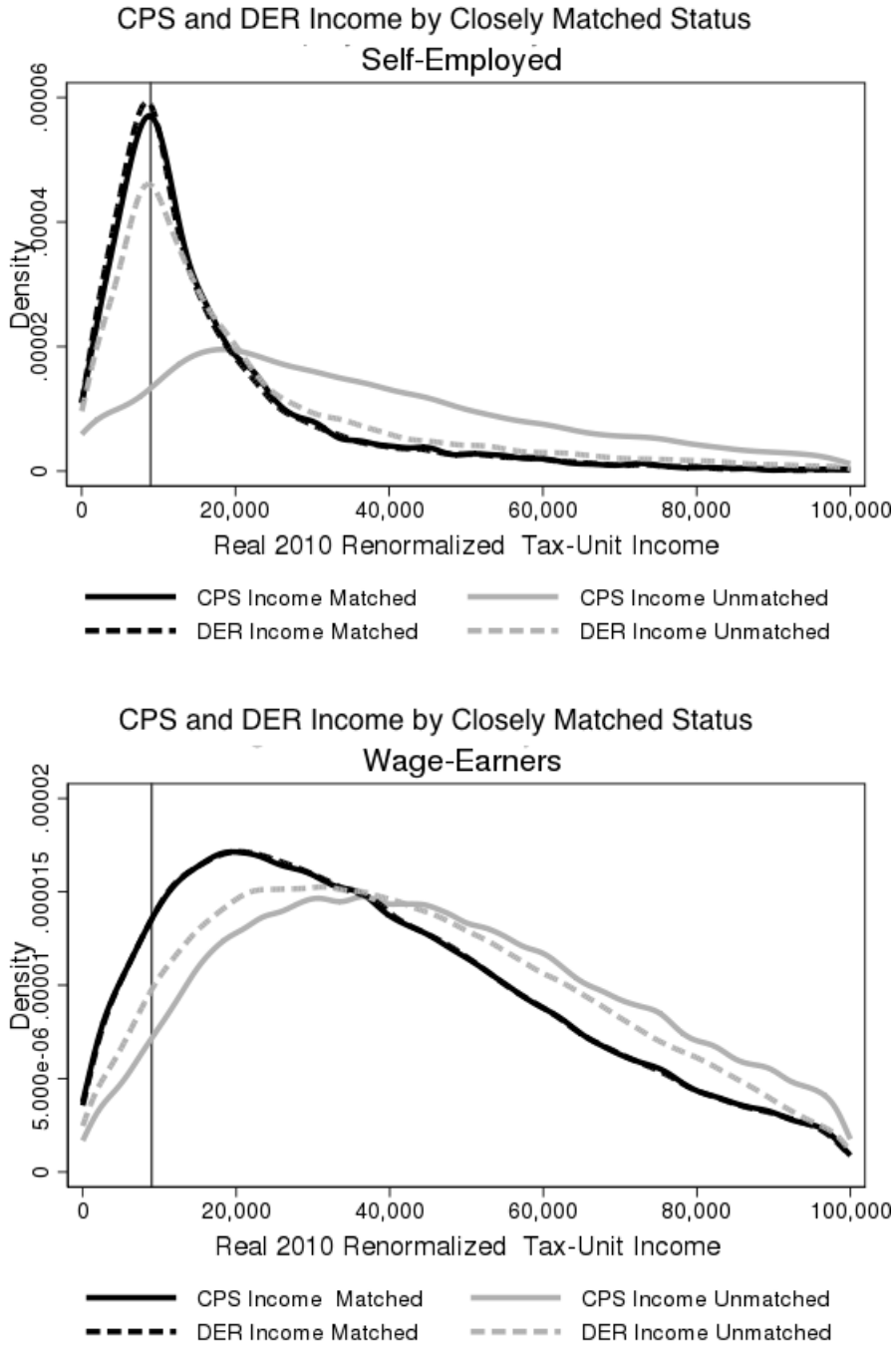
The top left figure repeats the top panel of Figure 1 with income renormalized so that all taxpayers have the same first EITC kink point as described in Section 3. For the self-employed average survey earnings are $13,000 higher than administrative earnings; for wage earners, the gap is only $3,000. The next three panels repeats the top panel of Figure 1 restricted to three unmarried samples (within each the first EITC kink is the same): 0 dependents, 1 dependent, and two or more dependents. For taxpayers without dependents, survey earnings are $13,000 higher than administrative earnings; for wage earners, the gap is only $1,000. For taxpayers one dependent, survey earnings are $9,000 higher than administrative earnings; for wage earners, the gap is only $2,000. For taxpayers two dependents, survey earnings are $7,000 higher than administrative earnings; for wage earners, the gap is only $2,000.

Figure 5: Reweighted Kernel Density Plot by Self-Employment Status



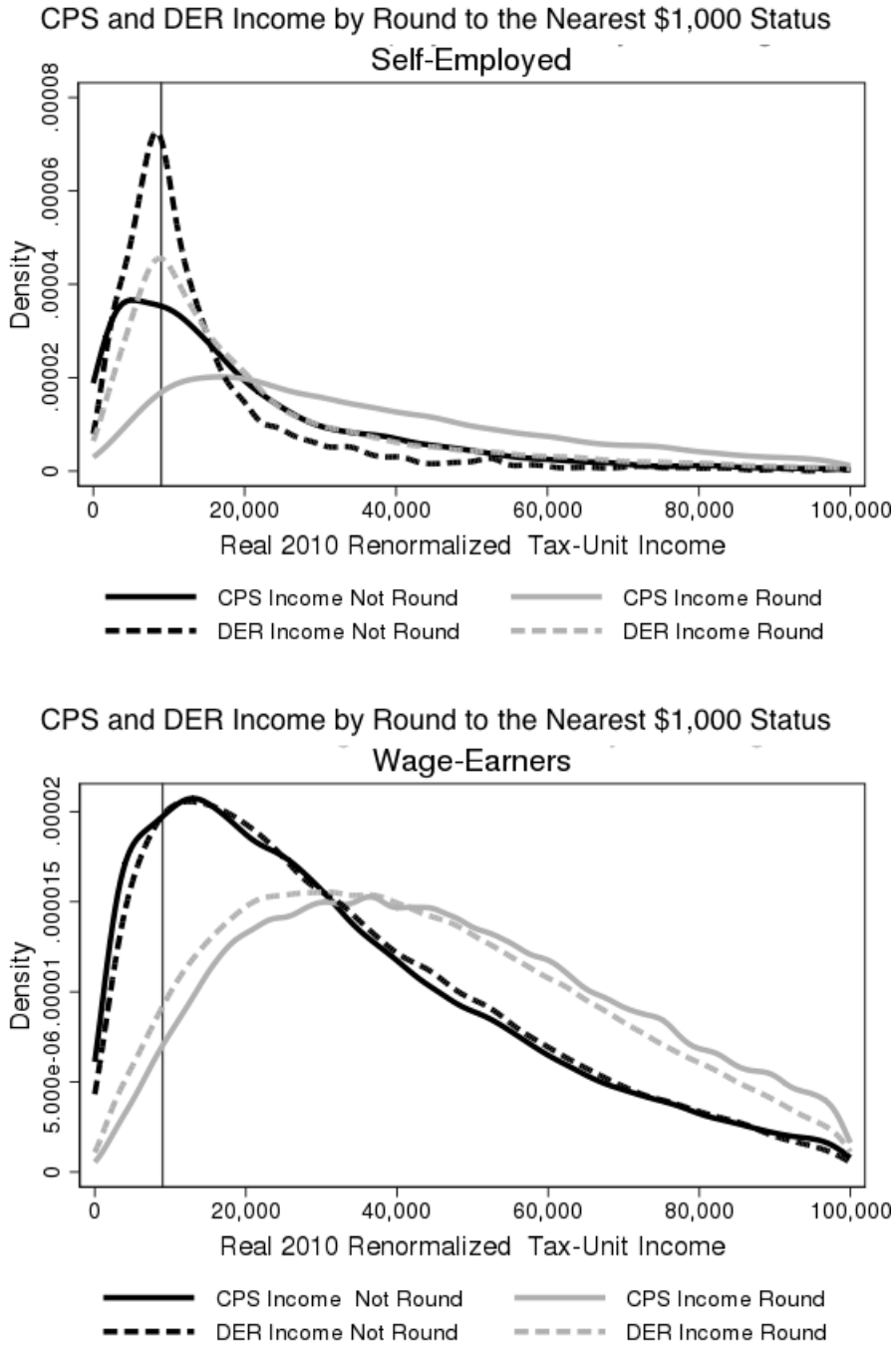
This figure repeats the top panel of Figure 1 with income renormalized so that all taxpayers have the same first EITC kink point and reweights the wage-earner distribution to be more like the self-employed distribution using IPW as described in Section 3.

Figure 6: Kernel Density Plots by Closely Matched Status



The top panel replicates the self-employed density plots in 4 separately for "closely matched" taxpayers (taxpayers whose survey (CPS) and administrative (DER) income are within $2,500) and "not closely matched" taxpayers. The bottom panel does the same for wage-earners. There is a 12.2 percent chance of being in the closely matched group for the self-employed and a 24.0 percent change for wage-earners. Survey and administrative income have a small gap that is not statistically different from zero. Survey earnings are $15,000 higher than administrative earnings for the not closely matched self-employed and $4,000 higher for not closely matched wage-earners.

Figure 7: Kernel Density Plots by Rounding Status



The top panel replicates the self-employed density plots in 4 separately for taxpayers that round and don't round to the nearest $1,000. The bottom panel does the same for wage-earners. This figure restricts income to be at least $1,000 so that both densities begin at the same place. For self-employed, survey earnings are $15,000 higher than administrative earnings if round to the nearest $1,000; for self-employed that are not round, the gap is $4,000. For wage-earners, survey earnings are $3,000 higher than administrative earnings if round to the nearest $1,000; for self-employed that are not round, the gap is -$1,000.