Reconciling Survey and Administrative Measures of Self-Employment

Katharine G. Abraham, *University of Maryland and National Bureau of Economic Research (NBER)*

John C. Haltiwanger, *University of Maryland and NBER*

Claire Hou, *University of Maryland and US Census Bureau*

Kristin Sandusky, *US Census Bureau*

James R. Spletzer, *US Census Bureau*

Good information on self-employment is needed to inform the ongoing discussion of the rise of the gig economy and its implications for workers. Tax data show significant growth in self-employment not captured in the Annual Social and Economic Supplement to the Current Population Survey (CPS-ASEC). The growing gap reflects both self-employment in tax data missing from the CPS-ASEC and self-employment misreported as wage and salary work. We document consistent patterns in the discrepancies between the tax and survey data but are able to explain only a modest share of the growing disagreement between them.

This paper is based on Katharine G. Abraham's Society of Labor Economists presidential address, delivered on May 3, 2019, at the society's annual meeting. John C. Haltiwanger was also a part-time Schedule A employee at the US Census Bureau at the time of the writing of this paper. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the view of the

Submitted February 28, 2020; Accepted October 12, 2020; Electronically published July 28, 2021.

Journal of Labor Economics, volume 39, number 4, October 2021.
 https://doi.org/10.1086/712187

I. Introduction

Changes in work arrangements argued to signal the breakdown of traditional employment relationships have attracted growing popular attention. Much of the focus has been on so-called gig workers, who provide services to businesses and households as independent contractors or freelancers. In recent years, mobile apps and online platforms have created new avenues for people willing to provide their services on a freelance basis to locate potential customers, contributing to interest in this sort of work arrangement. Similar to other self-employed individuals, gig workers do not enjoy the legal rights and protections afforded under federal minimum wage and overtime pay regulations, the unemployment insurance (UI) system, the workers' compensation system, and other laws and regulations written with more traditional employment arrangements in mind (Harris and Krueger 2015). Furthermore, those who rely primarily or exclusively on self-employment are markedly less likely to have health insurance or a retirement plan (Jackson, Looney, and Romnath 2017) and may have hours and earnings that are substantially more variable and less predictable. Responding to some of these concerns, in September 2019 the California legislature passed legislation limiting the circumstances under which firms may treat workers as independent contractors rather than employees. Proposition 22, passed in November 2020, partially reversed the law, specifying that app-based drivers should be treated as independent contractors (see, e.g., Conger 2020). This back and forth highlights the ongoing tension about worker classification in the era of online platform work.

Those expressing the most serious concerns about gig employment arrangements often appear to have in mind people who rely on such work as their primary source of income. In many cases, however, gig work supplements income from other sources rather than representing a person's primary income source. Some of the concerns that are especially relevant when a person relies

US Census Bureau. We have benefitted from the helpful comments of participants in the 2016 Society of Labor Economists annual meeting, the CRIW Conference on Measuring and Accounting for Innovation in the 21st Century, the Institute of Labor Economics (IZA) Labor Statistics Workshop on the Changing Structure of Work, the IZA Workshop on Labor Productivity and the Digital Economy, the 2018 and 2020 Labor and Employment Relations Association winter meetings, and seminars at the Federal Deposit Insurance Corporation, the University of Michigan, and the US Census Bureau. Eleanor Choi, Erica McEntarfer, and Lawrence Mishel provided valuable comments on drafts of the paper. The authors are grateful for financial support from the Russell Sage Foundation (award 85-17-06). All results have been reviewed to ensure that no confidential information has been disclosed (DRB-B0046-CED-20190425, CBDRB-FY20-CED006-0008, CBDRB-FY20-364, and CBDRB-FY20-415). Contact the corresponding author, Katharine Abraham, at kabraham@umd.edu. Information concerning access to the data used in this paper is available as supplemental material online.

primarily on gig work, such as the lack of protection under federal minimum wage and overtime rules, UI coverage, and employer-provided health insurance, arguably are less so when gig work supplements income from a primary job or other income source.

Good data on how work arrangements are changing are a prerequisite for evaluating the implications of these changes for US workers and their families as well as for the economy more broadly. Unfortunately, different sources of data send conflicting messages regarding recent trends. Policy makers and others interested in the evolution of work arrangements often turn to household survey data, where gig workers should be counted among the self-employed. In the Current Population Survey (CPS) and other household surveys, the percentage of the workforce that is self-employed has shown no upward trend and in fact has been drifting downward since at least the mid-1990s. In contrast, administrative data derived from tax filings support the popular perception that nonemployee work arrangements are a growing phenomenon (Katz and Krueger 2019a; Abraham et al. 2021).

In this paper, we compare self-employment status based on responses regarding prior year work activity in the Annual Social and Economic Supplement to the CPS (the CPS-ASEC) and self-employment status for the same people for the same year based on their income tax returns. The disagreement between these two data sources is both large and growing. Over the period from 1996 through 2015, 51.5% of those reporting CPS-ASEC self-employment income had no self-employment income for the same year on their tax returns. Even more striking, over the same period 66.7% of those with self-employment income on their tax returns did not report it in the CPS-ASEC. Furthermore, the number of people with self-employment income on their tax return but not in the CPS-ASEC grew from 7.8 million to 14.0 million between 1996 and 2015. As a share of those with labor earnings on their tax return, this is a 2.9 percentage point increase, from 5.7% to 8.6%. The growth includes people with self-employment income missing from the CPS-ASEC as well as people with self-employment income misreported as wage and salary income. People with both wage and salary income and self-employment income in the tax data but only wage and salary income in the CPS-ASEC—that is, people who fail to report a self-employment second job in the household survey data—account for the largest share of the growth. In contrast to the sizable growth in the number of people reporting self-employment earnings on their tax returns but not in the CPS-ASEC, there has been little growth in the number of people with self-employment earnings in the CPS-ASEC but not in the tax data.

Our examination of the situations in which self-employment is reported in tax data but not the CPS-ASEC suggests multiple contributing factors. For people who do not think of themselves as workers, even if they sometimes do things to earn money, prior year self-employment may not be especially salient and, as a result, not reported to the CPS interviewer. Consistent with

this explanation, among those with self-employment in the tax data, people who describe themselves at the time of the CPS interview as currently out of the labor force because of retirement or disability are notably overrepresented among those not reporting any prior year employment to the CPS interviewer. Similarly, people who have a wage and salary job that is their primary occupation but occasionally earn self-employment income on the side may not think to report it. Consistent with this story, those with lower or more transient self-employment earnings in the tax data are overrepresented in the group who fail to report a prior year self-employment second job in the CPS-ASEC. In contrast, workers with greater and more continuous self-employment earnings in the tax data are overrepresented in the group who misreport their self-employment income as wage and salary income, perhaps because for them self-employment seems more like a regular job. Although there are clear patterns in the characteristics of those with self-employment in the tax data but no CPS-ASEC self-employment, changes in these characteristics over the years covered by our data explain only a modest share of the growth in the discrepancy in reported self-employment in the two data sources.

Although the number of people with self-employment in the CPS-ASEC that is not reported in tax data has grown little in absolute size and has been stable as a share of all CPS-ASEC workers, there are a sizable number of such people. Tax returns are less likely to include self-employment income reported in the CPS-ASEC by people whose current primary activity is something other than work and/or who have a second CPS-ASEC self-employment job that generates relatively modest earnings.

II. Background and Literature Review

As discussed by Jackson, Looney, and Ramnath (2017), Abraham et al. (2021), and Abraham, Hershbein, and Houseman (2021), self-employment encompasses a variety of work arrangements. Some self-employed individuals run a business in which multiple factors of production are organized to provide products or services to customers. Others make money by selling their labor services as independent contractors, independent consultants, freelancers, platform workers, day laborers, or gig workers.

Both household surveys and tax data provide information on self-employment income. Household survey data commonly distinguish between incorporated and unincorporated self-employment. Labor earnings from incorporated self-employment are paid to business owners as wages or salaries. In tax data, these payments cannot be distinguished from other wage and salary income. In comparing reports of self-employment in household survey data to those in tax data, we therefore focus on the unincorporated self-employed.

Both of these two data sources are likely to miss some self-employment activity. Some respondents to surveys such as the CPS or the American

Community Survey (ACS) may fail to mention work they do not think of as a job. In addition, respondents who work for a business as an independent contractor or freelancer may be coded as employees rather than as self-employed. Administrative data capture only income that is reported to the tax authorities.

Several recent studies have produced evidence that the standard questions about current labor market activity asked in household surveys may miss some self-employment work. Abraham and Amaya (2019) report on a survey experiment administered in the summer of 2016 to a sample of respondents recruited via Amazon's Mechanical Turk. Study participants were asked the standard CPS employment questions for each member of their household. Questions probing for informal work done during the survey reference week raised the share of the relevant sample counted as employed by several percentage points and the share of the employed holding multiple jobs by 15–20 percentage points. Katz and Krueger (2019b) report that in a different sample of Amazon Mechanical Turk respondents surveyed in 2015, probing about small paid jobs after respondents had answered questions similar to those asked in the monthly CPS raised the multiple job holding rate from 39% to 77%. Bracha and Burke (2019) asked respondents to the 2015 Survey of Informal Work Participation, a survey administered to household heads participating in an online panel, a set of questions about informal work followed by the standard CPS employment questions. Incorporating informal work raised the overall employment rate by 4.5 percentage points and the multiple job holding rate by more than 11 percentage points compared with the estimates based on the standard CPS employment questions. Although it is important to be cautious about conclusions based on nonrepresentative samples or online panels, these findings nonetheless suggest that the standard CPS employment questions likely do not capture all self-employment work activity.

Results reported by Abraham, Hershbein, and Houseman (2021) based on a module of extra questions added to an ongoing Gallup telephone survey fielded in 2018 and 2019 suggest that miscategorization of self-employment as wage and salary work also may be a problem in household survey data. The standard Gallup question about work for an employer reads, "Thinking about your work situation over the past 7 days, have you been employed by an employer—even minimally like for an hour or more—from whom you receive money or goods?" Anyone answering yes to this question is coded as an employee. When asked a follow-up question, however, about 8% of those initially categorized as an employee on their main job indicated that they in fact were an independent contractor, independent consultant, or freelancer. Put differently, more than a quarter of those whose main job was self-employment initially reported themselves to be employees. The questions about employment in the CPS and the ACS differ from the Gallup question but may be vulnerable to similar response problems.

Although the CPS does not normally probe about the nature of reported employment arrangements, the periodic Contingent Worker Supplement (CWS) to the CPS asks about independent contract work. For anyone coded as an employee in the basic monthly CPS, the CWS asks, "Last week, (were/was) (you/NAME) working as an independent contractor, an independent consultant, or a free-lance worker? [That is, someone who obtains customers on their own to provide a product or service.]" About 14% of those the May 2017 CWS identified as independent contractors on their main job had been coded as an employee on that job in the basic monthly CPS (Abraham, Hershbein, and Houseman 2019). Because the bracketed clause in the CWS independent contractor question may have led some who really were independent contractors to answer no, this seems likely to be an underestimate of the extent of miscoding.

The evidence just described pertains to survey responses to questions about respondents' current labor market activity, but questions about sources of income over the previous calendar year (as asked on the CPS-ASEC) or previous 12 months (as asked on the ACS) likely suffer from similar problems. Roemer (2002) analyzed CPS-ASEC micro data for calendar years 1990, 1993, and 1996 linked to corresponding tax records supplied to the Census Bureau by the Social Security Administration. Focusing on wage earners, he found that 2%–3% of those reporting only wage income in the CPS-ASEC had only self-employment income in the tax records. Nicholas and Wiseman (2009) report similar results for 2003. Although we know of no prior studies showing that household survey reporting of self-employment income has become less accurate over time, there is considerable evidence that this has been the case for household survey reports of other types of income (see, e.g., Meyer, Mok, and Sullivan 2015).

Tax data have some important advantages compared with household survey data, but a significant limitation is that they capture only income reported to the tax authorities. In principle, tax returns should capture almost all self-employment income. Anyone with net income from unincorporated self-employment of $433 or more during the calendar year is required to file a tax return and report it on Schedule SE, the form used to calculate self-employment taxes. Tax filers with income from unincorporated self-employment also are required to document their business profits and losses either on Schedule C (sole proprietors) or on Schedule E (partners).

In practice, underreporting of self-employment income is a well-known problem. Whereas wage payments to an individual that exceed $600 during a calendar year are subject to information reporting on a Form W-2 submitted by the payer, self-employment income often has no associated information reporting. Businesses that pay nonemployee compensation of $600 or more to an independent contractor during the year are required to report it to the Internal Revenue Service (IRS) on a Form 1099-MISC, but this requirement does not apply to non-business-related payments. Since 2011, credit

card companies and payment processing companies, including platform companies, have been required to report certain payments to the IRS on a Form 1099-K. Credit card companies must report payments of any amount, but reporting by platform companies of payments to individuals who obtain work through the platform are required only for platform workers receiving more than $20,000 and more than 200 payments during the year. Even when a Form 1099-MISC or Form 1099-K is filed, it contains only gross payment amounts. To determine net self-employment earnings, expenses incurred in connection with that work also must be gauged (Government Accountability Office 2007; Slemrod et al. 2017). Not surprisingly, tax audit studies have shown that individual taxpayers report the large majority of wage and salary income but a notably smaller share of net nonfarm sole proprietor income and net farm income (Slemrod and Bakija 2008).

Another concern with using tax data to measure self-employment is that changes in filing incentives or reporting requirements may affect the comparability of the data over time. One development that could have affected the reporting of self-employment income is growing awareness of the Earned Income Tax Credit (EITC), a refundable tax credit calculated on the basis of the earnings of low-income tax filers (Chetty, Friedman, and Saez 2013; Collins et al. 2019; Mortenson and Whitten 2020). The recent introduction of the requirement that credit card companies and other payment processors report certain payments they handle to the IRS also could have affected filing behavior (Slemrod et al. 2017). Whether and to what extent such effects exist is, of course, an empirical question.

III. Data and Measurement

Our empirical analysis makes use of a data file containing records from the CPS-ASEC linked to tax information for the same individuals from the Detailed Earnings Record (DER) files supplied by the Social Security Administration to the Census Bureau. The CPS-ASEC, an annual supplement to the CPS, collects information on income and work activity during the prior calendar year. We have incorporated responses from the 1997 through the 2016 CPS-ASEC, which provide information on income and work activity for calendar years 1996–2015. We categorize a person as self-employed in the CPS-ASEC if her longest job during the year was unincorporated self-employment in which she had positive self-employment income or, for a person whose longest job was something other than unincorporated self-employed, if she had positive self-employment income from some other job. The CPS-ASEC questions about earnings other than on the longest job do not distinguish between incorporated and unincorporated self-employment, but information from the monthly CPS suggests that self-employment on a job other than a primary job is most likely to take the form of unincorporated

self-employment.[1] The questions about self-employment income on the CPS-ASEC ask the respondent to report earnings net of expenses.[2] In principle, the CPS-ASEC data should capture both primary and secondary self-employment.

CPS-ASEC information is imputed for people who responded to the basic monthly CPS portion of the survey but not to the supplement. Imputed values are used in constructing published CPS-ASEC statistics, and for consistency with the published estimates, we have retained the imputed records in our analysis. Dropping imputed values slightly reduces the disagreement in self-employment status between the CPS-ASEC and the tax records in the DER but does not materially affect the conclusions drawn from analysis of the linked data.

The DER is extracted from the Social Security Administration's Master Earnings File. It includes information on any earnings from unincorporated self-employment reported by taxpayers on a Schedule SE, together with information on the wage and salary earnings reported on any Form W-2s. Similar to the CPS-ASEC, the self-employment earnings reported for tax purposes are earnings net of expenses. We have DER earnings records for each CPS-ASEC respondent in our sample for whom there is a Protected Identification Key (PIK), an internal Census Bureau identifier based on the individual's encrypted Social Security number (SSN).[3]

Previous research based on linked CPS-ASEC and DER records, reported in Abraham et al. (2021), made use of an earlier version of the DER containing data for respondents to the 1997–2013 CPS-ASECs, from whom the CPS interviewers collected data for calendar years 1996–2012. Here we make use of an updated file that adds data for respondents to the 2014–16 CPS-ASECs, reporting for calendar years 2013–15. As discussed below, the current file also incorporates late returns for the final years included in the original file, so that the DER self-employment counts for 2011 and 2012 based on the updated file are slightly higher than the counts based on the earlier data. Because

[1] In data for the outgoing rotation groups interviewed in March 2003, 2007, and 2012, only about 2%–3% of those counted as self-employed in the monthly CPS due to either unincorporated self-employment on a first job or any self-employment on a second job are included because of incorporated self-employment on a second job.

[2] The CPS-ASEC question about self-employment earnings reads, "What were (name's/your) net earnings from this business/farm after expenses during (YEAR)?"

[3] Through 2005, the Census Bureau asked CPS-ASEC respondents who agreed to have their responses linked to administrative records for their SSNs. Starting in 2006, SSNs have not been collected, and respondents who do not want their records linked must opt out. The algorithm used in the Person Identification Validation System for record linkage when an SSN is not provided uses Felligi-Sunter agreement indices based on the individual's age, name, gender, and residence location. See Wagner and Layne (2014) for details.

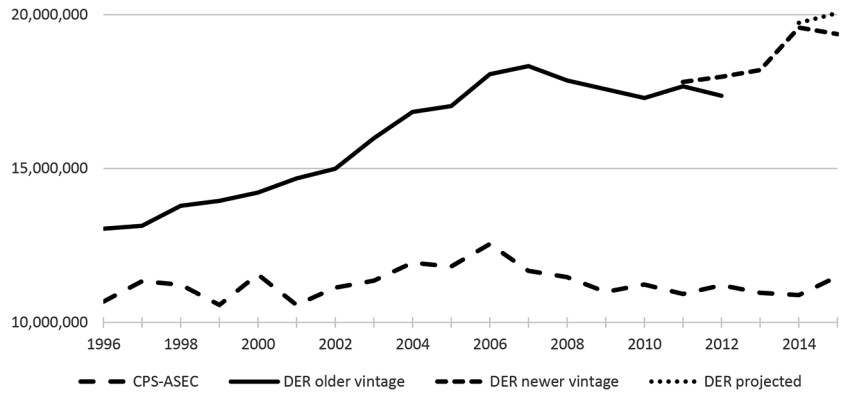

FIG. 1.—Estimated number of people with Current Population Survey Annual Social and Economic Supplement (CPS-ASEC) and Detailed Earnings Record (DER) self-employment earnings, 1996–2015. The data file includes CPS-ASEC records for individuals who have a Protected Identification Key linked to DER records for the same individuals. The solid line shows older vintage DER self-employment estimates, the dashed line shows current vintage DER estimates, and the dotted line shows projected DER values that incorporate adjustment for anticipated late returns. Tabulations are weighted. Source: Authors' tabulations of linked CPS-ASEC-DER data file.

estimates based on the earlier file are already in the public domain, however, releasing a full set of revised estimates based on the newer file would create complicated secondary disclosure issues. Other than the revised topside DER self-employment estimates shown in figures 1 and 2, the results we present therefore make use of DER records from the earlier data file for 1996–2012 self-employment and DER records from the newer data file for 2013–15 self-employment. Beyond slightly affecting the estimated self-employment counts for 2011 and 2012, using the original rather than the updated data has no material effect on our findings.[4]

Using the PIK, we link records from the CPS-ASEC to the administrative records from the DER for the years 1996–2015. The PIK is missing for

[4] In most of our analysis we use only the information from the DER on earnings in the year for which sample members report in the CPS-ASEC. The earlier version of the DER contained information for all CPS-ASEC respondents interviewed between 1997 and 2013 on their DER earnings in each year from 1978 through 2013; the later version contained information for all CPS-ASEC respondents interviewed between 1997 and 2016 on their DER earnings in each year from 1978 through 2016. The regressions reported in table 3 use information on prior year and subsequent year DER self-employment to assess the transiency of current year DER self-employment.

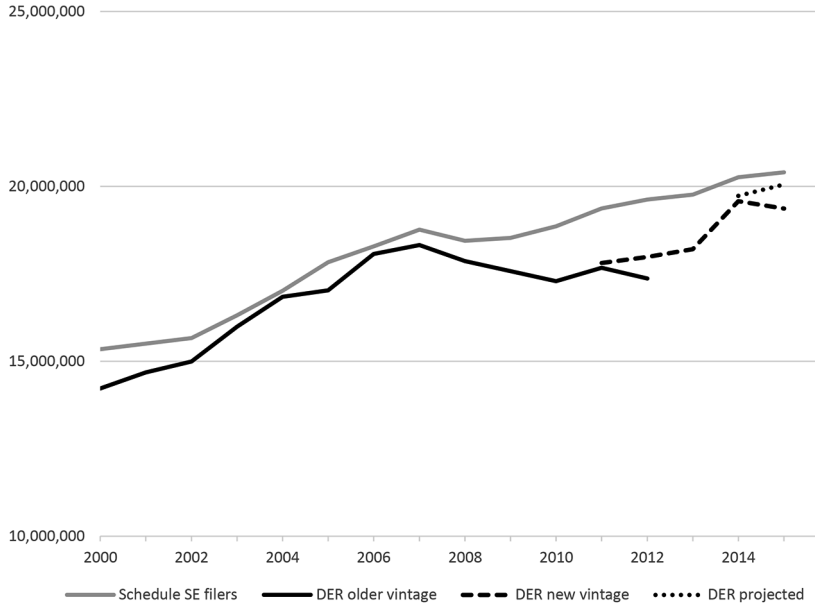

FIG. 2.—Estimated Detailed Earnings Record (DER) self-employment and population of Schedule SE filers. Schedule SE filers are population figures based on Internal Revenue Service data. DER estimates are based on weighted tabulations of Current Population Survey Annual Social and Economic Supplement (CPS-ASEC) responses for individuals who have a Protected Identification Key linked to DER data for the same individuals, as further described in the figure 1 legend. Source: Schedule SE filers from Collins et al. (2019); remaining series from authors' tabulations of linked CPS-ASEC-DER data file.

approximately 20%–30% of CPS-ASEC records, depending on the year.[5] We use propensity score methods to reweight the sample of people for whom we have a PIK so that their characteristics match those of the population as a whole. For each year, we regress an indicator for having a PIK on indicators for age group, gender, race, education, marital status, foreign-born status, state of residence, and whether the person reported being employed in the relevant CPS-ASEC. We then apply the coefficients from these models to calculate each individual's probability of having a PIK and apply a weight adjustment factor equal to the inverse of this probability to the CPS-ASEC estimation weight for those in the PIK sample.

[5] The adoption of an opt-out approach to consent for linkage in 2006 increased the share of records with a PIK (web appendix to Bollinger et al. 2019).

IV. Trends in Self-Employment in Household Survey and Administrative Data

We begin our investigation by examining the trends in estimated CPS-ASEC and DER self-employment as measured in the linked data file just described. One concern about using tax data to measure trends in self-employment is that changes in tax filing behavior could affect the comparability of estimates over time. As best as we can tell, however, this is not a major concern for estimates of the prevalence of self-employment in the DER over the period we study.

A. Trends in Estimated Self-Employment in the CPS-ASEC and the DER

Figure 1 plots annual estimates from the CPS-ASEC and the DER of the number of people with self-employment earnings for each year from 1996 through 2015. As already mentioned, a complication with administrative data such as those in the DER is that the incorporation of late-filed tax returns may lead to revisions in the numbers. The solid line segment in the upper part of figure 1 displays DER estimates for 1996–2012 based on the older version of the DER for which Abraham et al. (2021) reported selected estimates, and the dashed extension of that line displays 2011–15 values based on the updated DER file. The additional tax returns included in the later file raised the 2011 number only slightly but increased the estimated number of self-employed individuals in 2012 from 17.4 million to 18.0 million, or about 3.4%. The time interval between the end of 2015 and receipt of the current DER is similar to the time interval between the end of 2012 and receipt of the earlier file. Assuming that the time lags associated with the filing and processing of tax returns have not changed appreciably in recent years, we would expect the 2014 and 2015 self-employment estimates to be revised upward as late returns are added. The dotted extension of the DER line shows values for 2014 and 2015 that have been adjusted upward by factors equal to the ratios of the values for 2011 and 2012, respectively, from the updated file to those in the original file. Although we cannot be certain how late returns will affect the final DER self-employment numbers for 2014 and 2015, the dotted extension is an informed guess.

As shown clearly in the figure, in every year there are substantially more people with self-employment earnings in the DER than in the CPS-ASEC. In addition, whereas self-employment in the CPS-ASEC has been relatively stable, estimated DER self-employment grew considerably between 1996 and 2015, rising from 13.0 million in 1996 to 19.4 million in 2015, a number that our projection of late returns suggests may rise to perhaps 20.1 million. As a share of people with labor earnings on their tax returns, the currently available figures show an increase from 9.5% to 11.9%, a number that will rise somewhat if, as expected, late returns raise estimated 2015 DER self-employment.

Figure 2 compares the DER self-employment estimates shown in figure 1 to the total number of Schedule SE filers based on IRS records reported by Collins et al. (2019) for the period from 2000 to 2015. As we have described, the DER self-employment estimates were constructed by reweighting the sample of people in the CPS-ASEC who have a PIK to represent the population as a whole. Although the sample estimates diverge somewhat from the population figures for several years beginning in 2008 before moving back into closer alignment, the estimated change in DER self-employment for the 2000–2015 period as a whole is very similar to that in the IRS population counts. In what follows, we establish a common frame of reference for the administrative and household survey data by using the same weighted linked sample to measure the long-run change in estimated DER self-employment compared with estimated CPS-ASEC self-employment.

B. Tax-Filing Behavior and the Comparability of the DER Self-Employment Estimates over Time

A natural question is whether and to what extent changes in tax filing behavior might have contributed to the upward trend in estimated DER self-employment. Analyzing IRS data, Collins et al. (2019) find that over the period from 2007 to 2016, filers who claimed the EITC accounted for all of the increase in the number of people with Schedule SE self-employment income as a share of the tax workforce.[6] In our data, however, more than 45% of the growth in DER self-employment between 1996 and 2015 occurred among individuals with earnings of at least $50,000 (in 2015 dollars), suggesting that EITC claiming cannot be the whole story for the longer-run trend we observe. That said, although the changes in the EITC program since the mid-1990s have been modest (Crandall-Hollick 2018),[7] taxpayers could have become more knowledgeable over time about the EITC and how it operates, leading to changes in filing behavior. On its own, however, overall growth in self-employment among EITC claimants does not imply that responses to the tax incentives associated with the EITC have distorted the actual trend in DER self-employment.

For each dollar earned up to a specified level by workers in a tax filing unit (worker or worker plus spouse), the EITC offers a percentage credit

[6] Collins et al. (2019) define the tax workforce as tax filers with wage income, nonemployee compensation reported on a Form 1099-MISC, online platform income reported on a Form 1099-K, or Schedule SE income, plus nonfilers with wage income. This is more inclusive than the set of people with W-2 or Schedule SE income.

[7] Legislation passed in 2001 raised the income threshold at which married couples' EITC benefits begin to phase out. The American Recovery and Reinvestment Act of 2009 further extended the income range over which married couples are eligible for maximum benefits and introduced a higher credit for families with three or more children. For the reason explained in the next paragraph, only the last of these changes would have created a marginal incentive to report self-employment income.

that depends on the number of qualifying dependents in the unit. After reaching a maximum level, benefits remain constant over a specified income range and then fall by a percentage of each additional dollar earned until they are reduced to zero. Reporting self-employment income in order to claim the EITC is financially advantageous, at least to a point, for workers with qualifying dependents whose earned income (or, if married, whose earned income together with that of their spouse) otherwise lies below the first kink point of the applicable EITC schedule. For EITC eligibles with no dependent children or wage and salary incomes that place them above the schedule's first kink point, reporting self-employment income raises net tax liability and thus is financially disadvantageous.[8]

Given the structure of the incentives created by the EITC, we ask whether there has been disproportionate growth in reported self-employment among workers with qualifying dependents and own earned income (or own plus spouse's earned income) below the first kink point in the applicable EITC schedule. Qualifying dependents are children under age 19 (or under age 24 if enrolled full time in school) and may be claimed by parents, grandparents, or certain other relatives, provided the dependent resided with the person claiming the credit at least half the year. We do not have information on EITC claiming in our data, but from the CPS we know whether sample members are married and whether they have own children or grandchildren living with them who are either under age 19 or age 19–23 but enrolled in school. From the DER, we have information on the amounts of W-2 and Schedule SE income earned by the sample member and, if applicable, their spouse. Along with information from the CPS-ASEC on families' unearned income, this gives us a reasonable basis for identifying the group of interest.[9]

We use our linked data to estimate, by year, the number of EITC-eligible earners with qualifying dependents for whom the tax unit's earned income as

[8] The EITC pays tax filers with qualifying children who are below the EITC schedule's first kink point in the credit's phase-in range 34 cents per dollar earned if they have one child, 40 cents per dollar if they have two children, and (since 2009) 45 cents per dollar if they have three or more children, considerably more than the 15.30 cents in payroll tax liability associated with an extra dollar of self-employment income. In contrast, even in the phase-in range, the EITC pays only 7.65 cents on each extra dollar earned by eligible childless tax filers. For all tax filers, past the phase-in range, reporting additional self-employment income increases payroll tax liability but either has no effect on the EITC benefit amount (for those along the EITC schedule's plateau) or reduces it (for those past the EITC schedule's second kink point).

[9] We assume that children either under age 19 or age 19–23 and enrolled in school who live with a parent are qualifying dependents of the parent and his or her spouse, if present. If no parent is present in a family with such children but the CPS reference person is a grandparent of the children, we assume the children are qualifying dependents of the grandparent and his or her spouse, if present. Tax filing units with more than a modest amount of unearned income cannot claim the EITC. We use information on families' unearned income as reported in the CPS-ASEC to identify those not eligible for this reason.

recorded in the DER (wage and salary income plus self-employment income) places them along the phase-in portion of the EITC schedule. We also estimate the number of people in that group whose DER earnings include self-employment income. The total number of earners with qualifying dependents on the phase-in portion of the EITC schedule fell slightly between 1996 and 2015, but the share of that group reporting DER self-employment earnings rose from 11.1% to 20.6%. This is a much larger increase than the increase from 9.5% to 11.9% over the same period in the share of all DER earners with DER self-employment earnings. Had the share of people with qualifying dependents and earned incomes along the phase-in portion of the EITC schedule who reported DER self-employment earnings remained constant at the 11.6% observed in 1996, we estimate that overall DER self-employment would have been between 384,000 and 432,000 lower in 2015.[10] This represents 6.1% to 6.8% of the 6.3 million increase in DER self-employment actually observed between 1996 and 2015 and accounts for about 10% of the 2.4 percentage point increase in DER self-employment as a share of all DER employment over the same period. These numbers are consistent both with growing awareness of the EITC having caused strategic reporting of self-employment income to rise and with the effect on the overall number of DER self-employed being relatively small.

These estimates assume that all of the growth in the share of earners with qualifying dependents located on the phase-in portion of the EITC schedule who have self-employment income can be attributed to an increase in strategic reporting. Even among those for whom the EITC creates a net tax benefit to reporting self-employment income, learning about the credit could have had real labor supply effects as opposed to purely reporting effects, meaning that our estimates of the reporting effect could be too high. On the other hand, our estimates assume that people who are reporting strategically do not report so much additional self-employment income that they end up on the plateau of the EITC schedule. This could occur in some cases, but the fact that it is financially disadvantageous to report any more self-employment income than needed to reach the first kink point in the EITC schedule should discourage a purely strategic reporter from doing so. Finally, our estimates assume that our weighted linked CPS-ASEC-DER sample represents the population well and that we have correctly assigned dependents to the earners entitled to claim them for EITC purposes.[11] As a partial point of comparison, our estimates

[10] The smaller number is calculated assuming all of the added self-employed were people with reported wage and salary income; the larger is calculated assuming they were people who would not have reported any earned income absent an interest in claiming the EITC.

[11] An additional issue arises from the fact that in a small share of cases, one spouse in a family has a PIK but the other does not. In such cases, the spouse is not included in our sample, and we thus may not have an accurate measure of the tax filing unit's earned income.

of the total number of people with qualifying dependents who were EITC eligible in 2005 are close to those reported by Plueger (2009). Overall, the estimates just described seem most consistent with any trend in financially motivated reporting of DER self-employment earnings associated with growing awareness of the EITC having had a relatively small effect on overall DER self-employment.

Another recent change that potentially could have affected the reporting of self-employment income to the tax authorities is the introduction in 2011 of Form 1099-K, the information return that credit card and other payment processing entities are required to file to report the payments they process. This form provides the IRS with information on the credit card payments received by self-employed individuals. The estimates reported by Slemrod et al. (2017) suggest, however, that although receiving a Form 1099-K led some previously nonfiling businesses to submit a Schedule C, the new reporting requirement had only a minimal effect on the overall number of Schedule C filers. In 2011, only about 5% of Schedule C filers received a Form 1099-K. Of these, about 10% reported receipts within 5% of the amount on the form, of which as many as 30% may have filed a tax return because of the new information reporting. Taken together, the numbers suggest that the introduction of Form 1099-K reporting may have caused a one-time increase in the number of reported Schedule C businesses of perhaps 0.1%–0.2%.[12]

V. Disagreement between CPS-ASEC and DER Self-Employment Status

In addition to supporting comparisons of aggregate self-employment in the CPS-ASEC and DER, our linked data allow us to examine the agreement and disagreement between the information in the two data sources at the individual level. We are especially interested in how the number of people with self-employment income in one but not the other has changed over time. Even if the number of people with missing self-employment is large, the missing income associated with that self-employment might be relatively small. To see whether that is the case, we also look at the average earnings associated with missing self-employment.

A. Discrepancies in Self-Employment Status

Table 1 reports annual average estimates for each of four groups based on pooled data for the 1996–2015 period—people with no self-employment income in either the CPS-ASEC or the DER, people with self-employment

[12] A change that could have a larger effect on filing behavior is the recent decision of some large platform companies that previously had issued Form 1099-Ks to everyone working on the platform, even though not legally required to do so, to discontinue that practice (Collins et al. 2019). Because this change occurred after the end of our sample period, it would not have affected our results.

Table 1
Discrepancies in Estimates of Self-Employment Status Based on Current Population Survey Annual Social and Economic Supplement (CPS-ASEC) versus Detailed Earnings Record (DER) Earnings, 1996–2015

	Not Self-Employed in DER	Self-Employed in DER	Total
Not self-employed in CPS-ASEC:			
Number	205,849,371	10,978,424	216,827,794
Row share (%)	94.9	5.1	100.0
Column share (%)	97.3	66.7	95.1
Self-employed in CPS-ASEC:			
Number	5,808,202	5,471,298	11,279,501
Row share (%)	51.5	48.5	100.0
Column share (%)	2.7	33.3	4.9
Total:			
Number	211,657,573	16,449,722	228,107,295
Row share (%)	92.8	7.2	100.0
Column share (%)	100.0	100.0	100.0

SOURCE.—Authors' tabulations of linked CPS-ASEC-DER data file.
NOTE.—The data file includes CPS-ASEC records for individuals who have a Protected Identification Key linked to DER records for the same individuals. Numbers in the top row of each panel are 1996–2015 averages. Percentages are calculated from those averages. Tabulations are weighted.

income in the CPS-ASEC but not the DER, people with self-employment income in the DER but not the CPS-ASEC, and people with self-employment income in both the CPS-ASEC and the DER. Our primary interest lies with the "off-diagonals" in the table—the people with self-employment income in the CPS-ASEC but not the DER ({CPS-ASEC SE = 1, DER SE = 0}, the middle cell in the left column) and the people with self-employment income in the DER but not the CPS-ASEC ({CPS-ASEC SE = 0, DER SE = 1}, the middle cell in the top row).

There is a great deal of disagreement between self-employment as measured in the CPS-ASEC and the DER for the 1996–2015 period. Average self-employment in the DER over these 20 years (an estimated 16.4 million) is 45% larger than average self-employment in the CPS-ASEC (an estimated 11.3 million). Over the full 20-year period, among those with self-employment income in the CPS-ASEC, just over half (51.5%) have no self-employment income in the DER. The size of the other off-diagonal is even larger—over the full 20-year period, of those with self-employment income in the DER, fully two thirds (66.7%) do not have self-employment income in the CPS-ASEC.

That there is disagreement between the household survey and administrative data employment measures is not surprising, but the off-diagonal cells in the cross tabulations of self-employment status in the CPS-ASEC versus the DER are proportionally much larger than the off-diagonal cells in similar cross tabulations of wage and salary employment status. Abraham et al. (2013) found that over the 1996–2003 period, about 6% of individuals

who had UI earnings during the first quarter of the year reported no CPS wage and salary employment in a UI-covered sector during the year's first 3 months. Conversely, the same study found, about 18% of individuals reporting CPS wage and salary employment in a UI-covered sector during the first 3 months of the year had no first-quarter UI earnings. In our linked CPS-ASEC-DER data file for the 1996–2015 period, 9.5% of those with reported DER wage and salary income in a year had no CPS-ASEC wage and salary income in the same year, while 12.4% of those with CPS-ASEC wage and salary income had no reported DER wage and salary income.

Figure 3 examines the evolution of off-diagonals from the cross tabulation of the linked CPS-ASEC and DER data over time. Growth in the overall discrepancy between DER and CPS-ASEC self-employment could reflect either increases in the number of people who are self-employed in the DER but not the CPS-ASEC or decreases in the number of people who are self-employed in the CPS-ASEC but not the DER. For each year from 1996 through 2015, the figure plots the number of people who are self-employed in both the CPS-ASEC and the DER, together with the number of people in the {CPS-ASEC SE = 0, DER SE = 1} off-diagonal and the number of people in the {CPS-ASEC SE = 1, DER SE = 0} off-diagonal. The estimates were constructed using weighted micro data from the linked file. The figure shows that the number of people who are self-employed in the DER but not the CPS-ASEC grew substantially between 1996 and 2015, rising from

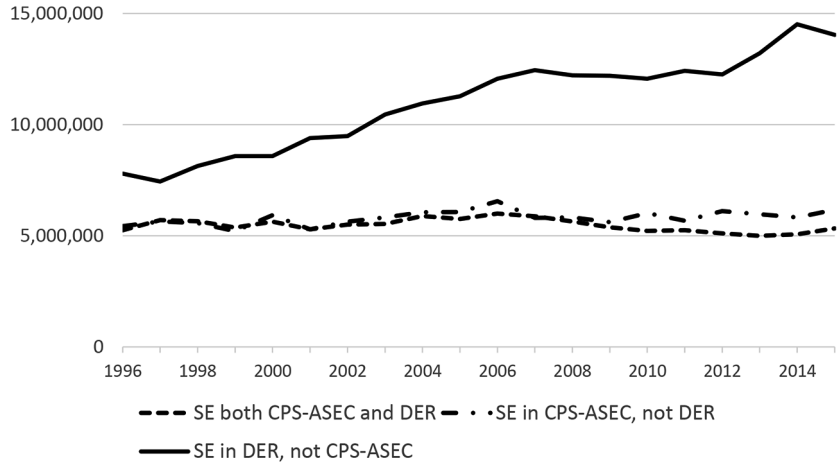

FIG. 3.—Estimated number of people in cells of Current Population Survey Annual Social and Economic Supplement (CPS-ASEC) versus Detailed Earnings Record (DER) self-employment earnings cross tabulations, by year, 1996–2015. The data file includes CPS-ASEC records for individuals who have a Protected Identification Key linked to DER records for the same individuals. Tabulations are weighted. SE = self-employment. Source: Authors' tabulations of linked CPS-ASEC-DER data file.

7.8 million to 14.0 million, or from 59.8% to 72.5% of the DER self-employed. As a share of those with earnings on their tax return, this represents a 2.9 percentage point increase, from 5.7% to 8.6% of all DER employed. In contrast, during the same period the number of people who are self-employed in the CPS-ASEC but not in the DER changed relatively little, rising by about 0.7 million—from around 5.4 million to around 6.2 million, or from 50.8% to 53.6% of the CPS-ASEC self-employed—and remaining at a constant 3.8% share of those with CPS-ASEC earnings.

B. Decomposing the Discrepancies in Self-Employment Status

To this point, we have not attempted to differentiate among the various situations that could give rise to a person having self-employment income reported in one data source but not the other. Disagreements may occur either because self-employment earnings are missing from the second data source or because earnings categorized as self-employment income in the first source appear as wage and salary income in the second. To further explore the discrepancies between the CPS-ASEC and DER measures of self-employment, we have looked more closely at the {CPS-ASEC SE = 0, DER SE = 1} and {CPS-ASEC SE = 1, DER SE = 0} off-diagonals, breaking the set of people in each of these two off-diagonals into several smaller and more homogeneous groups.

We begin this exploration by dividing the rapidly growing {CPS-ASEC SE = 0, DER SE = 1} off-diagonal group into three mutually exclusive categories:

1. *Missing CPS-ASEC self-employment, first job*. No wage and salary or self-employment income in the CPS-ASEC; self-employment income in the DER.[13]
2. *Missing CPS-ASEC self-employment, second job*. Only wage and salary income in the CPS-ASEC; both wage and salary and self-employment income in the DER.
3. *CPS-ASEC wage and salary job, classification issue*. Only wage and salary income in the CPS-ASEC; only self-employment income in the DER.

Those in the first two groups are missing self-employment in the CPS-ASEC. This could happen because a person does not consider herself to be employed and simply does not think to report it. Someone who is currently retired, for example, might not think to report their prior year self-employment income when responding to a household survey and thus fall into the category we have labeled "Missing CPS-ASEC self-employment,

[13] Those in this category may have either just self-employment income or both self-employment and wage and salary income in the DER.

first job." Similarly, someone who has a primary wage and salary job but earns occasional outside self-employment income also might not think to report it, putting them in the category we have labeled "Missing CPS-ASEC self-employment, second job."[14] We would expect missing CPS-ASEC self-employment to be more common when the associated earnings are low.

A person working as an independent contractor or "1099 worker" (someone whose self-employment earnings are reported on a Form 1099-MISC or Form 1099-K) but reporting their income as wage and salary earnings in the CPS-ASEC would fall into the group we have labeled "CPS-ASEC wage and salary job, classification issue." Given growing concerns about worker misclassification (see, e.g., Leberstein 2012), this group may be for some purposes the most interesting of the three in the {CPS-ASEC SE = 0, DER SE = 1} off-diagonal. Although we have no way of saying whether these individuals legally should have been considered employees, their CPS-ASEC responses suggest that that is how they see themselves.

We also have grouped those in the {CPS-ASEC SE = 1, DER SE = 0} off-diagonal into three mutually exclusive categories that parallel those defined for the {CPS-ASEC SE = 0, DER SE = 1} off-diagonal:

4. *Missing DER self-employment, first job*. No wage and salary or self-employment income in the DER; self-employment income in the CPS-ASEC.[15]
5. *Missing DER self-employment, second job*. Only wage and salary income in the DER; both wage and salary income and self-employment income in the CPS-ASEC.
6. *DER wage and salary job, classification issue*. Only wage and salary income in the DER; only self-employment income in the CPS-ASEC.

There are several ways that someone could end up in either the "Missing DER self-employment, first job" group or the "Missing DER self-employment, second job" group. The most obvious is that the person has earned taxable self-employment income that is not reported on their tax return—that is, that she is working off the books. It is also possible, however, that some of these people have self-employment earnings that are too low to trigger tax-reporting requirements or have reported their self-employment income elsewhere on their tax return.[16] Owners of incorporated businesses are people

[14] A professor who receives honoraria for reviewing papers or giving talks would be one example of someone who might fall into this category.

[15] Those in this category may have either just self-employment income or both self-employment and wage and salary income in the CPS-ASEC.

[16] Collins et al. (2019) report that in 2016, of those in the tax workforce who received either a Form 1099-MISC or a Form 1099-K reporting online platform earnings, about 11% reported a positive amount on the "other income" line of their Form 1040. The same was true of only 4.3% of those with only W-2 earnings.

we might expect to find represented in the "DER wage and salary job, classification issue" group. Some business owners may have correctly reported wage and salary income on their tax returns but been categorized incorrectly in the CPS-ASEC as unincorporated sole proprietors with self-employment earnings.

Figure 4*A* shows the evolution of the three groups within the {CPS-ASEC SE = 0, DER SE = 1} category, and figure 4*B* shows the evolution of those within the {CPS-ASEC SE = 1, DER SE = 0} category. The size of all three {CPS-ASEC SE = 0, DER SE = 1} groups has grown steadily over the 1996–2015 period. In contrast, employment in the three {CPS-ASEC SE = 1, DER SE = 0} groups has changed very little. The increasing size of all three of the groups in the {CPS-ASEC SE = 0, DER SE = 1} off-diagonal suggests that both missing self-employment and misreporting of self-employment as wage and salary work are growing problems in the CPS-ASEC.

One way to summarize the information presented in figure 4 is to ask what share of the 1996–2015 increase in the discrepancy between the estimated number of people with DER versus CPS-ASEC self-employment income each of the different off-diagonal groups can explain. For this purpose, we have averaged the numbers for the two starting years and the two ending years, then calculated the overall change in the discrepancy between those averaged end points. Table 2 shows the percentages of the growth in the overall discrepancy accounted for by each of the six off-diagonal groups.

As was apparent from figure 3, the growing gap between the DER and CPS-ASEC estimates of self-employment is accounted for entirely by the growing number of people identified as self-employed in the DER but not in the CPS-ASEC. The largest contributor to the increase in the gap is growth in the number of people who have both wage and salary income and self-employment income in the DER but only a wage and salary job in the CPS-ASEC. We refer to this group as missing a self-employment second job in the CPS-ASEC. Growth in the number of people we suspect are misclassified in the CPS-ASEC and growth in the number of people who have self-employment in the DER but no employment in the CPS-ASEC (i.e., people who are missing a self-employment first job in the CPS-ASEC) also have made sizable contributions. Taken together, these three groups more than explain the increasing gap between CPS-ASEC and DER self-employment. The net effect of changes in the size of the {CPS-ASEC SE = 1, DER SE = 0} off-diagonal subgroups is small and works in the wrong direction. Two of the three {CPS-ASEC SE = 1, DER SE = 0} subgroups have grown slightly, partially offsetting the contributions of growth in the size of the {CPS-ASEC SE = 0, DER SE = 1} off-diagonal subgroups to the size of the gap between DER and CPS-ASEC self-employment.

C. Self-Employment Earnings in the Off-Diagonals

A natural question about the marked growth in self-employment activity in the DER that is not captured in the CPS-ASEC is whether the associated

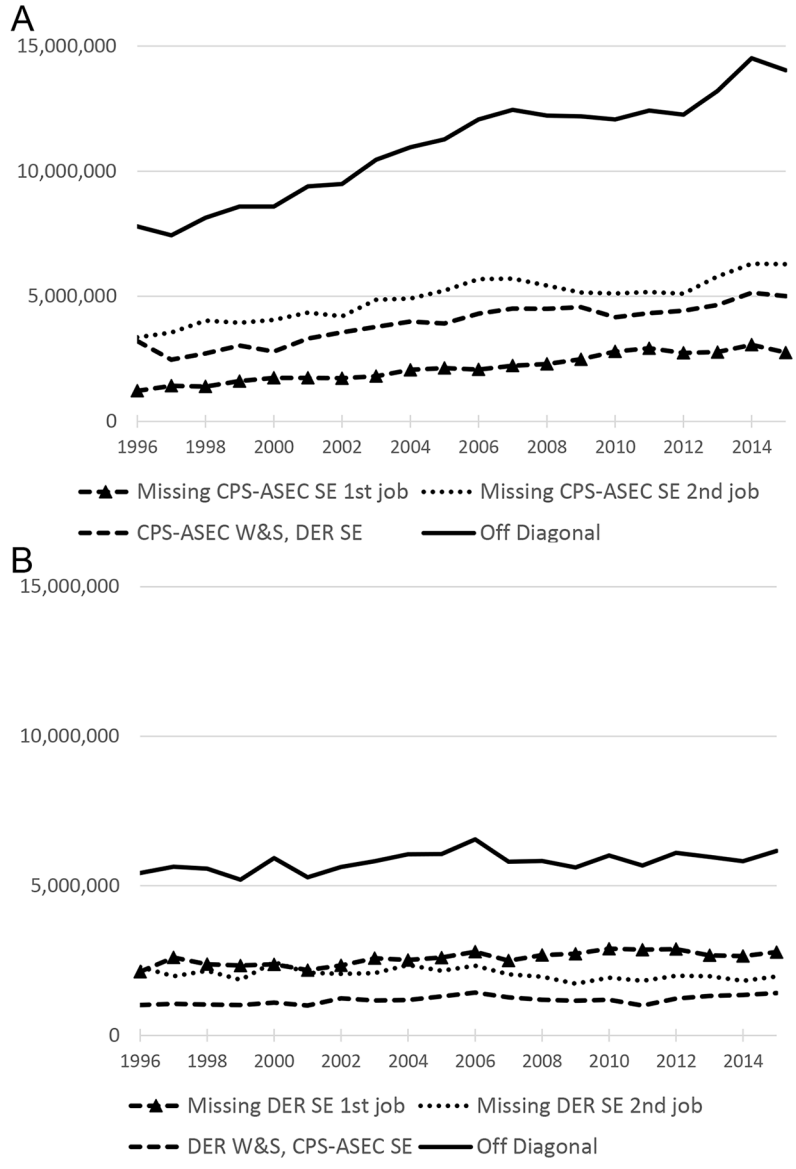

FIG. 4.—*A*, Disaggregated groups with Detailed Earnings Record (DER) but no Current Population Survey Annual Social and Economic Supplement (CPS-ASEC) self-employment earnings, 1996–2015. *B*, Disaggregated groups with CPS-ASEC but no DER self-employment earnings, 1996–2015. The data file includes CPS-ASEC records for individuals who have a Protected Identification Key linked to DER records for the same individuals. Tabulations are weighted. SE = self-employment; W&S = wage and salary employment. Source: Authors' tabulations of linked CPS-ASEC-DER data file.

Table 2
Percent Contributions to Growth in Discrepancy between Detailed Earnings Record (DER) and Current Population Survey Annual Social and Economic Supplement (CPS-ASEC) Self-Employment, 1996–97 to 2014–15

Off-Diagonal Category	Percentage of Growth in Discrepancy Explained
CPS-ASEC{SE = 0}, DER{SE = 1}:	
1. No CPS-ASEC employment	25.6
2. Self-employment second job not reported in CPS-ASEC	45.7
3. CPS-ASEC job misreported as wage and salary	36.1
CPS-ASEC{SE = 1}, DER{SE = 0}:	
4. No DER employment	−5.7
5. Self-employment second job not reported in DER	3.9
6. CPS-ASEC job misreported as self-employment	−5.6

SOURCE.—Authors' tabulations of linked CPS-ASEC-DER data file.
NOTE.—The data file includes CPS-ASEC records for individuals who have a Protected Identification Key linked to DER records for the same individuals. Reported percentages are the share of growth in discrepancy between the numbers of people with self-employment earnings in DER and CPS-ASEC from 1996–97 to 2014–15 explained by the change in size of each listed category. Tabulations are weighted.

earnings are economically significant. If the increase in the number of people with self-employment income in the DER but missing from the CPS-ASEC were due primarily to gig workers earning very small amounts of money, for example, its importance would be less obvious. A similar question can be asked about the self-employment reported in the CPS-ASEC but missing from the DER. In the latter case, a significant amount of self-employment income not being reported to the tax authorities could suggest serious tax compliance concerns.

Our linked data file provides DER self-employment earnings for everyone with {DER SE = 1} and CPS-ASEC self-employment earnings for everyone with {CPS-ASEC SE = 1}. We are able to compare the earnings in the two data sources only for the people who are self-employed in both. Over the 1996–2015 period, earnings for this group averaged about $43,000 in the CPS-ASEC and about $24,500 in the DER (in 2015 dollars), a surprisingly large discrepancy. One possible explanation is that despite the instructions to report self-employment earnings net of expenses, some CPS-ASEC respondents may have reported their gross earnings. Another possibility is that in the CPS-ASEC, some respondents own up to self-employment income they have hidden from the tax authorities. Alternatively, some self-employed people could be overstating the income-related expenses they report to the IRS but reporting their net incomes more accurately in the CPS-ASEC.

Figure 5*A* and figure 5*B* plot average self-employment earnings for individuals in each of the three subgroups of the {CPS-ASEC SE = 0, DER SE = 1} off-diagonal and in each of the three subgroups of the {CPS-ASEC SE = 1, DER SE = 0} off-diagonal, respectively, from 1996 through 2015. In figure 5*A*, the off-diagonal group with the lowest DER self-employment

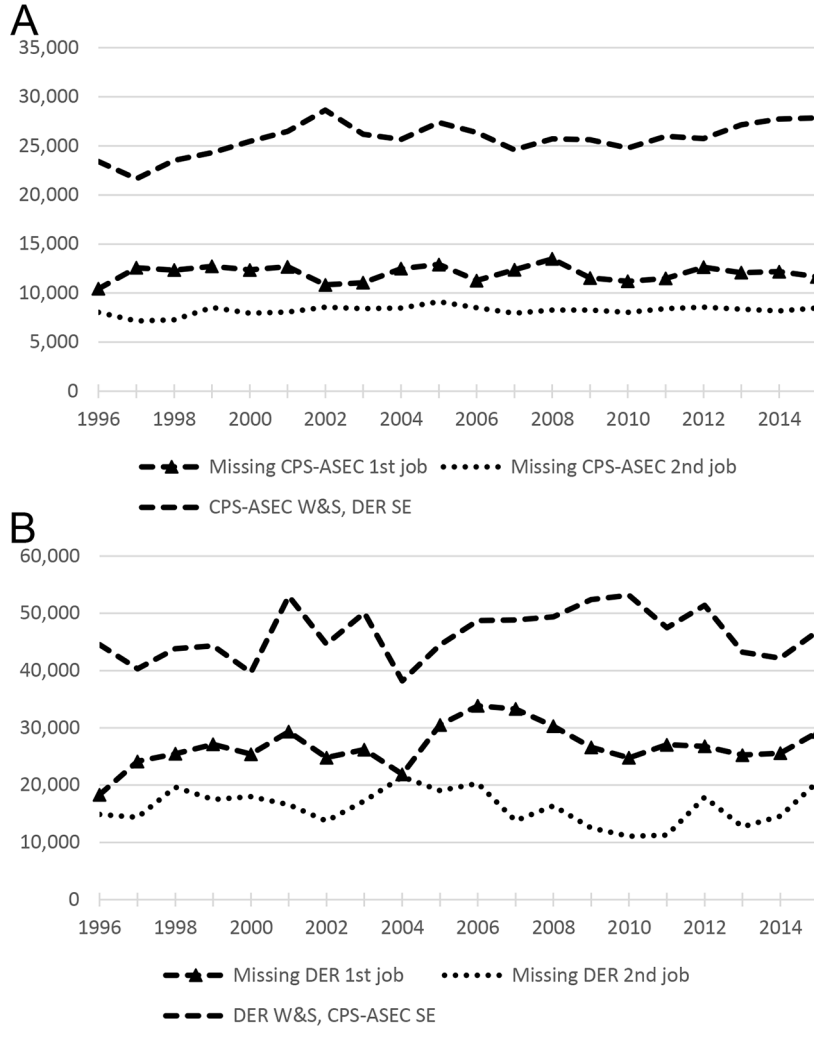

FIG. 5.—*A*, Self-employment earnings for disaggregated groups with Detailed Earnings Record (DER) but no Current Population Survey Annual Social and Economic Supplement (CPS-ASEC) self-employment, 2015 dollars, 1996–2015. *B*, Self-employment earnings for disaggregated groups with CPS-ASEC but no DER self-employment, 2015 dollars, 1996–2015. The data file includes CPS-ASEC records for individuals who have a Protected Identification Key linked to DER records for the same individuals. Tabulations are weighted. SE = self-employment; W&S = wage and salary employment. Source: Authors' tabulations of linked CPS-ASEC-DER data file.

earnings is the group with missing CPS-ASEC self-employment second jobs, but even for that group, DER self-employment earnings averaged over $8,000 (in 2015 dollars) over the 20-year period. DER self-employment earnings are higher, averaging about $12,000 (in 2015 dollars), for those with a missing CPS-ASEC self-employment first job. For the group of DER self-employed who appear to be misclassified as wage and salary workers in the CPS-ASEC, DER self-employment earnings averaged about $26,000 (in 2015 dollars). These are far from trivial amounts. Had the growth in the number of people in the {CPS-ASEC SE = 0, DER SE = 1} off-diagonal we have documented consisted disproportionately of people doing low-earnings work, we would have expected average earnings in some or all of these cells to fall. Instead, average DER self-employment earnings have been essentially flat for the two groups with missing CPS-ASEC self-employment and have trended upward for the DER self-employed group we suspect the CPS-ASEC has coded as wage and salary.

It is possible, of course, that the averages do not tell the whole story. Even if there had been growing numbers of low-earnings self-employed people in the DER who are not captured in the CPS-ASEC, their effect on average DER self-employment earnings could have been masked by increases in DER self-employment earnings higher in the distribution. In addition to looking at average earnings, we also have looked at earnings at the 10th, 25th, 50th, 75th, and 90th percentiles of the earnings distributions in each of the three {CPS-ASEC SE = 0, DER SE = 1} off-diagonal subgroups. None of these data series point to any different conclusion than the mean earnings series shown in figure 5*A*.

Another way to think about the economic significance of the DER self-employment earnings accounted for by people who are not self-employed in the CPS-ASEC is to look at their share of all DER self-employment earnings. Over the 1996–2015 period as a whole, 55% of DER self-employment earnings were reported by people with no self-employment income in the CPS-ASEC. Reflecting the substantial growth in the number of people in the {CPS-ASEC SE = 0, DER SE = 1} off-diagonal, this share rose from 40%–50% in the late 1990s to about 70% by the end of our time period.

For completeness, figure 5*B* reports average CPS-ASEC self-employment earnings for those in the three {CPS-ASEC SE = 1, DER SE = 0} off-diagonal subgroups. Because the size of the group with CPS-ASEC self-employment but no DER self-employment changed little over the 1996–2015 period, there is not the same a priori reason to be concerned that its composition has changed in any particular fashion. Reported CPS-ASEC self-employment earnings are sizable for each of the three subgroups; the sizable average earnings for the groups with a missing first or second DER self-employment job, in particular, raise potential tax compliance concerns. As with the earnings of those in the {CPS-ASEC SE = 0, DER SE = 1} off-diagonal and its subgroups, however, none of the series shows any obvious trend.

VI. Who Is in the Off-Diagonals?

There is clearly considerable disagreement between self-employment as measured in the CPS-ASEC versus the DER. To better understand the reasons for this disagreement, we have fit linear probability models to explore the characteristics associated with being in the {CPS-ASEC SE = 0, DER SE = 1} or {CPS-ASEC SE = 1, DER SE = 0} off-diagonals and their various subcomponents. The sample for the first set of regressions is everyone with DER self-employment income. In the four models fit for this sample, the dependent variable equals 1 if the person was in the off-diagonal or one of its three subcomponents and otherwise equals 0. In the second set of regressions, the sample is everyone with CPS-ASEC self-employment income. The dependent variables in those models capture whether the person was in the {CPS-ASEC SE = 1, DER SE = 0} off-diagonal or one of its three subcomponents. All of the regressions are fit using pooled data for the 1996–2015 period. The regression coefficients can be used in conjunction with changes in the values of the explanatory variables to assess whether changes in the observables help to explain observed changes in the sizes of the off-diagonals and their subcomponents.

As already noted, there are several reasons why a person might end up in the {CPS-ASEC SE = 0, DER SE = 1} off-diagonal. In the models reported in table 3, we include predictors intended to capture people whose primary activity at the time of the CPS interview is something other than work—dummy variables for currently being in school, out of the labor force and retired, or out of the labor force and disabled. Even if someone in these categories had self-employment income during the prior calendar year, she might be especially likely not to think to report it. A second set of predictors is intended to capture self-employment that generated little earnings or was transient. These include dummy variables for quartile of the DER self-employment earnings distribution based on the pooled sample, a dummy variable for the person's self-employment having either begun or ended during the calendar year, and a dummy variable for the person having self-employment in the current year but not the preceding or following year. We would expect those with higher DER self-employment earnings to be less likely and those with more transient DER self-employment to be more likely to have a missing CPS-ASEC first or second job. A third possible contributor to the {CPS-ASEC SE = 0, DER SE = 1} off-diagonal is that the respondent misreported independent contractor work as wage and salary employment in the CPS-ASEC. All else the same, we would expect self-employment that generates more income or is more persistent to seem more like a traditional job to the person doing it and thus be more likely to be misclassified in this way.

In addition to the key variables just described, the table 3 models control for education, race, foreign-born status, gender, marital status, presence of children in the family and the interaction between marital status and the presence

Table 3
Effects of Person and Job Characteristics on the Probability That a Person with Detailed Earnings Record (DER) Self-Employment Earnings Does Not Have Current Population Survey Annual Social and Economic Supplement (CPS-ASEC) Self-Employment Earnings

	Explanatory Variable Means (1)	CPS-ASEC SE = 0, DER SE = 1| DER SE = 1 (2)	Missing CPS-ASEC Self-Employment 1st Job| DER SE = 1 (3)	Missing CPS-ASEC Self-Employment 2nd Job| DER SE = 1 (4)	Misclassification| DER SE = 1 (5)
Age:					
15–24	.054	.112***	.035***	.065***	.012**
25–34	.179	.043***	.013***	.025***	.005
35–44	.245	Omitted	Omitted	Omitted	Omitted
45–54	.245	−.032***	−.002***	−.027***	−.004
55–64	.176	−.059***	.004***	−.068***	.005
≥65	.101	−.050***	.058***	−.149***	.041***
Education:					
Less than high school	.108	.023***	.038***	−.025***	.009**
High school graduate	.276	Omitted	Omitted	Omitted	Omitted
Some college	.256	−.015***	−.033***	.028***	−.010***
College graduate	.214	.012***	−.047***	.061***	−.003
Postcollege	.147	.038***	−.077***	.078***	.037***
Race:					
Black	.079	.102***	.058***	.041***	.003
White	.857	Omitted	Omitted	Omitted	Omitted
Other	.064	.010**	.017***	−.017***	.010**

Demographics:					
Foreign born	.192	.084***	.002***	−.027***	.110***
Male	.599	−.002	−.050***	.036***	.012***
Married	.671	−.011***	.007	−.011***	−.007***
Any children	.406	.037***	.043***	.014***	−.020***
Married × any children	.327	−.044***	−.032***	−.026***	.014***
Reporter:					
Proxy response	.439	.054***	.018***	.021***	.016***
Missing self/proxy	.037	.016	.013	.003	.000
Primary activity not work:					
Enrolled in school	.026	.030***	.130***	−.134***	.033***
NILF, retired	.056	.263***	.606***	−.239***	−.104***
NILF, disabled	.011	.197***	.592***	−.269***	−.126***
DER self-employment $:					
First quartile	.250	Omitted	Omitted	Omitted	Omitted
Second quartile	.250	−.067***	.002	−.108***	.039***
Third quartile	.250	−.121***	.017***	−.245***	.106***
Fourth quartile	.250	−.173***	−.002***	−.329***	.158***
Self-employment:					
Began or ended	.303	.100***	.012***	.132***	−.044***
Current year only	.163	.180***	.028***	.233***	−.082***
R^2		.122	.254	.204	.058
Dependent variable mean		.668	.115	.305	.232
Unweighted N		172,000	172,000	172,000	172,000

SOURCE.—Authors' analysis of linked CPS-ASEC-DER data file.

NOTE.—Regressions are based on pooled 1996–2015 data. The sample is all people with DER self-employment in the year of observation. Models also include year dummies. Regressions are weighted. NILF = not in labor force.

** $p < .05$.

*** $p < .01$.

of children. They also control for whether the person's information was reported by a proxy, on the grounds that proxy reporters may be more likely to provide an inaccurate report. In practice, it turns out, there are not large differences between self-reports and proxy reports. All of the models contain a full set of year dummies.[17]

The first column of table 3 shows the weighted mean values for all of the explanatory variables in the reported models exclusive of the year dummies. The remaining four columns show the coefficient estimates for the {CPS-ASEC SE = 0, DER SE = 1} off-diagonal model and each of the three subcomponents models. Because these are linear probability models and the three subcategories are mutually exclusive and exhaustive, the estimated coefficients in the third through fifth columns of the table sum to the corresponding coefficient in the second column. A positive coefficient on a variable implies that people or jobs with that characteristic are overrepresented in the off-diagonal or off-diagonal subgroup; a negative coefficient implies that they are underrepresented.

The coefficient estimates in the table's second column indicate that people with prior year DER self-employment who are currently out of the labor force and either retired or disabled are especially likely not to report prior year self-employment in the CPS-ASEC. Having low DER self-employment earnings or transient DER self-employment activity also makes it more likely that a person fails to report the DER self-employment in the CPS-ASEC. Because they may reflect offsetting effects on being in the different off-diagonal subgroups, however, these coefficients do not tell the whole story.

We turn next to the models for the three different subgroups of people in the {CPS-ASEC SE = 0, DER SE = 1} off-diagonal. Compared with others with DER self-employment income, people who are out of the labor force and say they are retired or disabled are especially overrepresented in the subgroup with a missing CPS-ASEC self-employment first job. In contrast, they are underrepresented among the subgroup with a missing CPS-ASEC self-employment second job and among the subgroup misreporting independent contractor work as a wage and salary job. To a lesser extent, students with DER self-employment income also are overrepresented among those with a missing CPS-ASEC self-employment first job and underrepresented among those with a missing CPS-ASEC self-employment second job. None of these groups accounts for an especially large share of the DER self-employed, but they are interesting because they are identifiable populations who seem especially likely not to think of themselves as workers. The foreign born are notably more

[17] Industry dummies were not included in the models because industry is not consistently available. In the DER, it is available only beginning in 2002, and even then only for self-employed sole proprietors, not for those with self-employment income from a partnership. In the CPS-ASEC, industry of self-employment is available only for those who were self-employed on their longest job. When we included industry dummies where available with a separate dummy for cases missing an industry, they added little to the models' explanatory power.

Table 4
Accounting for Aggregate Change in Share of People with Detailed Earnings Record (DER) Self-Employment Earnings Who Do Not Have Current Population Survey Annual Social and Economic Supplement (CPS-ASEC) Self-Employment Earnings, 1996–97 to 2014–15

	CPS-ASEC SE = 0, DER SE = 1| DER SE = 1	Missing CPS-ASEC Self-Employment 1st Job| DER SE = 1	Missing CPS-ASEC Self-Employment 2nd Job| DER SE = 1	Misclassification| DER SE = 1
ΔY, 1996–97 to 2014–15	.151	.048	.050	.053
$\Delta X\beta/\Delta Y$ (X = personal characteristics) (%)	12.4	21.6	27.8	−10.7
$\Delta X\beta/\Delta Y$ (X = DER SE job characteristics) (%)	3.6	1.7	−7.2	15.8
$\Delta X\beta/\Delta Y$ (X = personal and job characteristics) (%)	16.1	23.3	20.6	5.1

SOURCE.—Authors' analysis of linked CPS-ASEC-DER data file.

NOTE.—Calculations are based on coefficient estimates from table 3 combined with changes in mean characteristics of explanatory variables included in the regression between 1996–97 and 2014–15. Personal characteristics include dummy variables for age group, education group, race, foreign born, gender, marital status, presence of own children, interaction of marital status with presence of own children, self/proxy, school enrollment, not in the labor force and retired, and not in the labor force and disabled. DER self-employed job characteristics include self-employment earnings quartile and dummies for continuity of self-employment.

likely than others to misreport their self-employment work as a wage and salary job. Turning to the characteristics of DER self-employment work, those with DER self-employment that is more transient or generated lower earnings are overrepresented in the subgroup that is missing a CPS-ASEC self-employment second job but underrepresented in the subgroup misreporting independent contractor work as a wage and salary job. Putting this somewhat differently, the type of self-employment that is least likely to be missing from the CPS-ASEC reports—self-employment that generates relatively high and stable earnings—is exactly the type of self-employment that is most likely to be misreported in the CPS-ASEC as wage and salary work.

Using the estimates from table 3, we examine how much of the change in the proportion of the DER self-employed who are in the {CPS-ASEC SE = 0, DER SE = 1} off-diagonal or one of its subgroups can be accounted for by observable changes in the distribution of the characteristics of the DER self-employed over time. For these calculations, we focus on the long differences from 1996–97 to 2014–15 in these shares as shown in the first row of table 4. The next three rows of table 4 report the shares of the long differences accounted for by changes in the personal characteristic variables, changes in the DER self-employment variables, and changes in both sets of variables together, respectively.[18]

[18] For this purpose, personal characteristic variables include all of the explanatory variables shown in table 3 except for the dummy variables for DER self-employment earnings quartile, began or ended DER self-employment during the calendar year, and had DER self-employment in the current year but not the previous or subsequent year. The latter are what we term the DER self-employment variables.

Table 5
Effects of Person and Job Characteristics on the Probability That a Person with Current Population Survey Annual Social and Economic Supplement (CPS-ASEC) Self-Employment Earnings Does Not Have Detailed Earnings Record (DER) Self-Employment Earnings

	Explanatory Variable Means (1)	CPS-ASEC SE = 1, DER SE = 0| CPS-ASEC SE = 1 (2)	Missing DER Self-Employment 1st Job|CPS-ASEC SE = 1 (3)	Missing DER Self-Employment 2nd Job|CPS-ASEC SE = 1 (4)	Misclassification| CPS-ASEC SE = 1 (5)
Age:					
15–24	.052	.157***	−.050***	.143***	.065***
25–34	.159	.026***	−.020***	.029***	.017***
35–44	.247	Omitted	Omitted	Omitted	Omitted
45–54	.263	−.032***	.004	−.019***	−.017***
55–64	.188	−.063***	.025***	−.057***	−.030***
≥65	.090	−.034***	.135***	−.127***	−.042***
Education:					
Less than high school	.102	.043***	.101***	−.062***	.004
High school graduate	.272	Omitted	Omitted	Omitted	Omitted
Some college	.277	.012***	−.022***	.045***	−.011***
College graduate	.213	.011***	−.048***	.070***	−.011***
Postcollege	.137	−.018***	−.088***	.097***	−.027***
Race:					
Black	.069	.149***	.059***	.053***	.037***
White	.875	Omitted	Omitted	Omitted	Omitted
Other	.056	.043***	.007	.006	.029***

Demographics:					
Foreign born	.149	−.004	.006*	−.031***	.021***
Male	.606	.003	−.031***	.047***	−.013***
Married	.664	−.049***	−.033***	−.004	−.012***
Any children	.382	−.087***	−.045***	−.020***	−.022***
Married × any children	.317	.034***	.031***	−.001	.004
Reporter:					
Proxy response	.403	.022***	.000	−.001	.024***
Missing self/proxy	.046	.013	.015	−.030*	.028**
Primary activity not work:					
Enrolled in school	.026	.034***	.123***	−.064***	−.026***
NILF, retired	.022	.061***	.158***	−.142***	.046***
NILF, disabled	.006	.131***	.274***	−.160***	.017
CPS self-employment $:					
First quartile	.250	Omitted	Omitted	Omitted	Omitted
Second quartile	.250	−.213***	−.004	−.261***	.051***
Third quartile	.250	−.302***	−.021***	−.354***	.073***
Fourth quartile	.250	−.332***	−.052***	−.377***	.096***
R^2		.111	.045	.192	.022
Dependent variable mean		.514	.228	.181	.105
Unweighted N		118,000	118,000	118,000	118,000

SOURCE.—Authors' analysis of linked CPS-ASEC-DER data file.

NOTE.—Regressions are based on pooled 1996–2015 data. The sample is all people with CPS-ASEC self-employment in the year of observation. Models also include year dummies. Regressions are weighted. NILF = not in labor force.

* $p < .10$.

** $p < .05$.

*** $p < .01$.

The combined effects of all of the characteristics controlled for in the table 3 regressions account for about 16% of the increase from 1996–97 to 2014–15 in the share of the DER self-employed who are in the overall off-diagonal, mostly due to changes in the personal characteristic variables. As in table 3, however, looking at the overall off-diagonal does not tell the whole story. Changes in the personal and job characteristics of those with DER self-employment can explain about 23% of the long-run increase in the share who are missing a self-employment first job in the CPS-ASEC and about 21% of the long-run increase in the share who are missing a self-employment second job in the CPS-ASEC. In both cases, changes in personal characteristics do most of the work; for missing self-employment second jobs, changes in the DER self-employment variables actually work in the wrong direction. Changing characteristics account for only about 5% of the increase in the share of the DER self-employed who are classified as wage and salary workers in the CPS-ASEC.

Although the estimated coefficients in table 3 exhibit sensible patterns that are consistent with our hypotheses about who is most likely to be in the {CPS-ASEC SE = 0, DER SE = 1} off-diagonal, the R^2 values in these regressions are relatively low, and the changing distribution of observable characteristics accounts for only a modest fraction of the increase in the share of the DER self-employed who are in the off-diagonal. In short, much of the large and rapidly growing {CPS-ASEC SE = 0, DER SE = 1} off-diagonal is not explained by the factors we observe.

Table 5 contains results for the other off-diagonal, that is, people with CPS-ASEC self-employment who do not have DER self-employment. The first column in the table again reports the weighted mean values of the explanatory variables used in our models, exclusive of the year dummies. The coefficients estimates for the model explaining the overall {CPS-ASEC SE = 1, DER SE = 0} off-diagonal are shown in the second column and the coefficients estimates for the subgroup models are shown in the third through fifth columns.[19] Some of the determinants of being in the {CPS-ASEC SE = 1, DER SE = 0} off-diagonal are qualitatively similar to those for being in the {CPS-ASEC SE = 0, DER SE = 1} off-diagonal, although the effect sizes in the {CPS-ASEC SE = 1, DER SE = 0} models are generally smaller and the likely explanations for them are somewhat different. Among those reporting CPS-ASEC self-employment, students, self-described retirees, and self-described disabled individuals are somewhat more likely not to have DER self-employment income, but this is the net result of being more likely to have a missing DER self-employment first job and less likely to have a missing DER self-employment second job. Similarly, individuals with low CPS-ASEC self-employment earnings are more likely to have a missing DER self-employment second job but

[19] The regressions in table 5 are not fully comparable to those in table 3, as we do not have good measures of the transience of CPS-ASEC self-employment.

less likely to belong to the group for which self-employment in the CPS-ASEC appears as wage and salary work in the DER. Although these systematic patterns in the likelihood of appearing in the {CPS-ASEC SE = 1, DER SE = 0} off-diagonal and its subcomponents are interesting, the change over our time period in the share of the CPS-ASEC self-employed who have no DER self-employment is very small, and, as such, the results of calculations analogous to those shown in table 4 are neither interesting nor informative. We do not report them here.

VII. Conclusion

Because the nature of a person's employment arrangements can have such an important effect on her well-being, it is important to have good information about how those arrangements are changing. The distinction between wage and salary employment versus self-employment is especially important in this regard. Although some people prefer to be self-employed and others benefit from the opportunity to supplement their incomes with self-employment work, the self-employed lack many of the legal protections afforded to wage and salary workers, are not covered by employer-provided health insurance and retirement plans, and may experience greater volatility in their earnings. Policy makers typically rely on household survey data to understand ongoing changes in the labor market, but our results raise questions about how well these surveys are measuring self-employment.

We find that there is a growing gap between the prevalence of self-employment as captured in tax data (specifically, the DER) and their prevalence as measured for the same set of people in household survey data (specifically, the CPS-ASEC). Whereas the DER shows that there was substantial growth in self-employment over the 20 years from 1996 through 2015, no such growth can be seen in the CPS-ASEC data. One might be concerned that measures of the prevalence of self-employment based on tax data could have been distorted by changes in tax filing behavior, but as best we can tell this does not seem to have been an important factor over the period we study.

Although we believe they contain valuable information, we do not argue that tax data on their own are the answer to the challenge of developing reliable measures of self-employment activity. The tax data we analyze have captured growth in self-employment that is missing from the household survey data, but there is also self-employment income recorded in the household survey data that is not reported to the tax authorities. Rather, our argument is that the increasing amount of self-employment captured in tax data but not the CPS-ASEC should lead scholars to take a closer look at what those data—and household survey data on self-employment more generally—may be missing and why it is happening. The evidence we have presented on the correlates of self-employment reported to the tax authorities but not

measured in the household survey data may help to inform future research and experimentation.

One possible path forward may be to develop periodic surveys that probe more directly regarding respondents' employment arrangements. A lesson we take away from our exploration of the off-diagonals in reported self-employment—most especially, self-employment that is reported in the tax data but does not appear in the household survey data—is that some of the problem occurs among people who do not think of themselves primarily as workers. Related to this, people who do relatively small amounts of self-employment work on an irregular basis, whether as a first or second job, appear to be less likely than others with self-employment income to report it. This suggests that asking explicitly about informal or other work the respondent has not already reported as a job or business may produce better measures of work activity. Our analysis also suggests that another source of problems is that some self-employed people think of themselves as working for an employer and, as a result, are miscategorized as wage and salary workers. Again, probing questions included on periodic surveys could help to identify these situations.

Sources of data other than household survey data, including not only tax data but also private data generated by platform companies, financial institutions, and financial services companies, may help with monitoring the ways in which work and the income it generates are changing. Although no single source of data is likely to tell us everything we might like to know about self-employment and self-employment income, information from each source can provide a valuable check on the information from the others. Indeed, modeled estimates that incorporate information from multiple sources ultimately may provide the most reliable statistics. In an era in which anecdotal evidence suggests that the labor market may be changing in significant ways, taking steps to improve the measurement of self-employment and other aspects of the employment relationship would be a worthwhile investment.

References

Abraham, Katharine G., and Ashley Amaya. 2019. Probing for informal work activity. *Journal of Official Statistics* 35, no. 3:487–508.

Abraham, Katharine G., John C. Haltiwanger, Kristin Sandusky, and James R. Spletzer. 2013. Exploring differences in employment between household and establishment data. *Journal of Labor Economics* 31, no. S1:S129–S172.

———. 2021. Measuring the gig economy: Current knowledge and open issues. In *Measuring and accounting for innovation in the 21st century*, ed. Carol Corrado, Javier Miranda, Jonathan Haskel, and Dan Sichel. Chicago: University of Chicago Press.

Abraham, Katharine G., Brad Hershbein, and Susan N. Houseman. 2019. Independent contract, informal, and online intermediary work: Preliminary evidence on developing better measures in household surveys. Working paper, W. E. Upjohn Institute for Employment Research, Kalamazoo, MI.

———. 2021. Contract work at older ages. *Journal of Pension Economics and Finance* 20, no. 3:426–47.

Bollinger, Christopher R., Barry T. Hirsch, Charles M. Hokayem, and James P. Ziliak. 2019. Trouble in the tails? What we know about earnings nonresponse thirty years after Lillard, Smith, and Welch. *Journal of Political Economy* 127, no. 5:2143–85.

Bracha, Anat, and Mary A. Burke. 2019. How big is the gig? The extensive margin, the intensive margin, and the hidden margin. Working paper, Boston Federal Reserve Bank, Boston.

Chetty, Raj, John N. Friedman, and Emmanuel Saez. 2013. Using differences in knowledge across neighborhoods to uncover the impacts of the EITC on earnings. *American Economic Review* 103, no. 7:2683–721.

Collins, Brett, Andrew Garin, Emilie Jackson, Dmitri Koustas, and Mark Payne. 2019. Is gig work replacing traditional employment? Evidence from two decades of tax returns. Working paper, Statistics of Income, Internal Revenue Service, Washington, DC.

Conger, Kate. 2020. Uber and Lyft drivers in California will remain contractors. *New York Times*, November 7.

Crandall-Hollick, Margot L. 2018. The Earned Income Tax Credit (EITC): A brief legislative history. Congressional Research Service Report no. R44825.

Government Accountability Office. 2007. Tax gap: A strategy for reducing the gap should include options for addressing sole proprietor noncompliance. Report to the Committee on Finance, US Senate.

Harris, Seth D., and Alan B. Krueger. 2015. A proposal for modernizing labor laws for twenty-first century work: The "independent worker." Hamilton Project Discussion Paper no. 2015-10, Brookings Institution, Washington, DC.

Jackson, Emilie, Adam Looney, and Shanthi Ramnath. 2017. The rise of alternative work arrangements: Evidence and implications for tax filing and benefit coverage. Working Paper no. 114, Office of Tax Analysis, Washington, DC.

Katz, Lawrence F., and Alan B. Krueger. 2019a. The rise and nature of alternative work arrangements in the United States, 1995–2015. *ILR Review* 72, no. 2:382–416.

———. 2019b. Understanding trends in alternative work arrangements in the United States. *Russell Sage Foundation Journal of the Social Sciences* 5, no. 5:132–46.

Leberstein, Sarah. 2012. Independent contractor misclassification imposes huge costs on workers and federal and state treasuries. National Employment Law Project report.

Meyer, Bruce D., Wallace K. C. Mok, and James X. Sullivan. 2015. Household surveys in crisis. *Journal of Economic Perspectives* 29, no. 4:199–226.

Mortenson, Jacob A., and Andrew Whitten. 2020. Bunching to maximize tax credits: Evidence from kinks in the U.S. tax schedule. *American Economic Journal: Economic Policy* 12, no. 3:402–32.

Nicholas, Joyce, and Michael Wiseman. 2009. Elderly poverty and supplemental security income. *Social Security Bulletin* 69, no. 1:45–73.

Plueger, Dean. 2009. Earned Income Tax Credit participation rate for tax year 2005. In *Proceedings of the 2009 IRS Research Conference*, 151–95. Washington, DC: Internal Revenue Service.

Roemer, Marc. 2002. Using administrative earnings records to assess wage data quality in the March Current Population Survey and the Survey of Income and Program Participation. Working paper, US Census Bureau, Suitland, MD.

Slemrod, Joel, and Jon Bakija. 2008. *Taxing ourselves: A citizen's guide to the debate over taxes*. 4th ed. Cambridge, MA: MIT Press.

Slemrod, Joel, Brett Collins, Jeffrey Hoopes, Daniel Reck, and Michael Sebastiani. 2017. Does credit-card information reporting improve small-business tax compliance? *Journal of Public Economics* 149:1–19.

Wagner, Deborah, and Mary Layne. 2014. The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications' (CARRA) record linkage software. CARRA Working Paper no. 2014-01, US Census Bureau, Suitland, MD.