



Tulane Economics Working Paper Series

Using Public Information to Estimate Self-Employment Earnings of Informal Suppliers

James Alm
Department of Economics
Tulane University
jalm@tulane.edu

Brian Erard
B. Erard & Associates

Working Paper 1517
July 2015

Abstract

An enduring problem in the analysis of tax evasion is the difficulty of its measurement. An especially troublesome component of tax evasion arises from informal suppliers, such as self-employed domestic workers, street-side vendors, and moonlighting tradesmen. We develop in this paper a new approach for estimating self-employment earnings of informal suppliers. Our methodology involves using national survey results on self-employment earnings within a carefully selected set of industry categories where informal activities are believed to be concentrated. Then, by comparing these national survey results on self-employment earnings to Internal Revenue Service statistics on the amounts actually reported for tax purposes, it is possible to estimate the extent of noncompliance within the selected industry categories. Our methodology relies on survey respondents being reasonably forthcoming about their earnings, which we are able to confirm through some validation exercises.

Keywords: tax evasion, informal suppliers
JEL codes: H26, C81

USING PUBLIC INFORMATION TO ESTIMATE SELF-EMPLOYMENT EARNINGS OF INFORMAL SUPPLIERS

James Alm and Brian Erard*

An enduring problem in the analysis of tax evasion is the difficulty of its measurement. An especially troublesome component of tax evasion arises from informal suppliers, such as self-employed domestic workers, street-side vendors, and moonlighting tradesmen. We develop in this paper a new approach for estimating self-employment earnings of informal suppliers. Our methodology involves using national survey results on self-employment earnings within a carefully selected set of industry categories where informal activities are believed to be concentrated. Then, by comparing these national survey results on self-employment earnings to Internal Revenue Service statistics on the amounts actually reported for tax purposes, it is possible to estimate the extent of noncompliance within the selected industry categories. Our methodology relies on survey respondents being reasonably forthcoming about their earnings, which we are able to confirm through some validation exercises.

* Tulane University and B. Erard & Associates. This research was funded by the Internal Revenue Service (IRS) (TIRNO-03-P-00651). The views expressed are those of the authors and do not necessarily reflect the opinions of the IRS or of any researchers working within the IRS. We thank the IRS Office of Research Analysis and Statistics, and in particular, Charles Bennett, Kim Bloomquist, Ed Emblom, and Alan Plumley for their helpful comments and assistance with understanding and using the NRP data. We also thank three anonymous referees and the Editors for many useful suggestions. Please address all correspondence to James Alm, Department of Economics, 208 Tilton Hall, Tulane University, New Orleans, LA 70118 USA (phone +1 504 862 8344; fax +1 504 865 5869; email jalm@tulane.edu).

INTRODUCTION

An enduring problem in the analysis of tax evasion is the difficulty of its measurement, a problem that is deeply rooted in the obvious incentives for evaders to conceal their unlawful behavior. There have been many approaches to measurement, and researchers have become increasingly creative in devising new methods, often based on traces (or indirect indicators) of noncompliance. Even so, all of these methods are subject to criticism and, in many cases, a good deal of skepticism.¹ However, accurate measurement is crucial along multiple dimensions. For researchers, weak measures of tax compliance hamper their ability to credibly test theories of compliance behavior. In the case of policy-makers, these weaknesses limit their capacity to measure the true effects of government policies on the distribution of income and other economic indicators and to design appropriate economic policies. For tax administrators, compliance measurement problems frustrate their efforts to allocate services and enforcement activities where they are most needed.

An especially troublesome component of tax evasion arises from informal suppliers. The U.S. Internal Revenue Service (IRS) defines “informal suppliers” as “...individuals who provide products or services through informal arrangements which frequently involve cash-related transactions or ‘off the books’ accounting practice” (IRS, 1996a, 43). Examples include self-employed domestic workers, street-side vendors, and moonlighting tradesmen. Conceptually, the informal economy within which such individuals operate includes all types of market economic activity that are potentially under-measured in the National Income Accounts, owing to the vendors’ informal business styles (e.g., sales in cash, lack of adequate records of sales and purchases). More relevant for our purposes here is the extent to which legally taxable self-

employment earnings from informal market activities are reported – or not reported – on individual income tax returns; that is, what is the extent of underreporting among informal suppliers with legal sources of income?²

In this paper we develop a new approach to address this question. Our methodology involves using national survey results on self-employment earnings within a carefully selected set of industry categories where informal activities are believed to be concentrated. By focusing on a carefully chosen set of industry categories, we believe that the resulting estimate of informal sector income should encompass the vast majority of all earnings of informal suppliers as well as the earnings of formal suppliers within these categories. Then, by comparing these national survey estimates of actual self-employment earnings within the selected categories to statistics on the portion of these earnings ultimately reported for tax purposes in these same categories, it is possible to estimate the extent of noncompliance with respect to tax reporting requirements.

Importantly, our methodology for estimating informal supplier earnings does not rely on data sources that are outdated, costly, proprietary, or difficult to obtain. Rather, it exploits national survey results that are readily available to the public. We rely primarily on statistics derived from the Current Population Survey (CPS), but we also use of findings from the Consumer Expenditure Survey (CES) for one of our industry categories. Similar information is frequently available in many countries outside of the U.S. The main data requirement is tabulations of reported self-employment earnings or consumer expenditures within selected

¹ See Schneider and Enste (2002), Alm (2012), and Slemrod and Weber (2012) for detailed discussions and critiques of these many approaches.

² The difficulties posed by informal suppliers are well-recognized by the U.S. Department of the Treasury; for example, see Treasury Inspector General for Tax Administration (2013). For discussions of various methods used to estimate informal supplier income from legal sources of income, see McCrohan and Smith (1986), Smith and Adams (1987), McCrohan and Sugrue (1988), Luttikhuizen and Kazemier (2000), and Brown (2001, 2002). For developing countries, see the May 2011 special issue of *The Review of Income and Wealth* on measuring the informal economy in developing countries, especially the papers by Kulshreshtha (2011) and Joshi, Amoranto, and Hasan (2011). Useful surveys are in Losby et al. (2002) and OECD (2002). See also IRS (1996b).

industry categories based on a statistically representative survey.³ Our approach therefore offers considerable scope for measuring unreported income in other countries by using similar data sources that break down reported self-employment earnings by industry. In this respect, our approach has some parallels to attempts to measure tax evasion via public surveys, in which the relationship between expenditures and income is estimated for wage and salary workers and is then used to infer the true income of self-employed workers based upon their reported expenditures.⁴

As with all existing methods, there are some challenges with the implementation of our approach. First, although responses to CPS surveys are strictly confidential, some respondents may still be reluctant to reveal details on their earnings and sources of employment. For this reason, our methodology is likely to provide a lower bound on self-employment earnings within the selected industry categories. Second, although the CPS includes detailed information on earnings and employment, it is necessary to perform imputations to account for some respondents who misreport their self-employment earnings as wages, as well as to account for earnings of second job holders in relevant industries. Even so, we find that our estimates seem reasonable and also appear to outperform examiner-based estimates derived from intensive random audits of tax returns.

³ For example, see the publicly available microdata survey information available at The World Bank website <http://microdata.worldbank.org/index.php/catalog/central>.

⁴ For example, Pissarides and Weber (1989) estimate the relationship between expenditures and income for wage and salary workers in the United Kingdom, for whom reported income is thought to be an accurate measure of their “true income”, and they then use this estimated relationship to infer the true income of the self-employed based upon their reported expenditures. Ekici and Besim (2014) follow a similar procedure for North Cyprus; see also Hurst, Li, and Pugsley (2014) for the United States, Schuetze (2002) for Canada, Johansson (2005) for Finland, Gibson, Kim, and Chung (2009) for Korea and Russia, and Engstrom and Holmond (2009) for Sweden. Feldman and Slemrod (2007) use a comparable methodology but estimate instead for U.S. workers the itemized deductions-income relationship rather than the expenditures-income relationship. These studies typically find that the self-employed underreport by amounts that vary between 1/6 and 1/2 of their true income. In contrast to these studies, our focus is on unreported income within industries dominated by informal suppliers, rather than for all sole proprietors. As discussed later, when our methodology is extended to all sole proprietors, it implies underreporting rates ranging from 19 to 54 percent over a period of years.

Below we discuss previous approaches to measuring informal sector income, including their limitations. We then lay out our methodology, present our results, and conclude by discussing the relative advantages and disadvantages of our approach.

PREVIOUS MEASUREMENT APPROACHES FOR INFORMAL SUPPLIER INCOME

Owing in large part to the lack of a paper trail, tax evasion among informal suppliers can be especially difficult to uncover through examinations, even intensive large scale random audit programs such as the National Research Program (NRP) or its predecessor, the Taxpayer Compliance Measurement Program (TCMP).

As a result, the IRS has in the past attempted to address this potentially severe non-detection problem with the aid of supplementary information from a special survey that it periodically commissioned of consumer purchases in the informal sector, most recently by the University of Michigan Survey Research Center for 1985-1986. Under this approach, estimation of the extent of underreporting among informal suppliers followed a two-step process. The first step was to use the survey results to develop an estimate of the aggregate gross receipts of informal suppliers; this estimate of gross receipts was then converted into an estimate of net self-employment income using an assumption about the ratio of net earnings to gross receipts. In the second step, IRS researchers attempted to identify informal suppliers on the basis of the limited information available in the TCMP and to assess how much of this net income was actually reported on their tax returns. The difference between the estimated amount of what was actually earned by informal suppliers (based on the survey) and the amount that was ultimately reported on tax returns (based on the TCMP) served as the IRS measure of noncompliance.

The starting point in this estimation process therefore relied upon the use of a survey of purchases from informal suppliers that is now well out-of-date. However, even if an updated survey were available, the accuracy of the approach depends critically on one's ability to distinguish formal from informal transactions, not only on the survey but also on tax returns. Unfortunately, the kinds of information collected from these sources, especially from tax returns, do not seem adequate for this purpose.

An alternative approach that is used in various countries for estimating informal supplier income is to conduct a specialized employment survey.⁵ While similar to a standard labor force survey such as the CPS, these specialized surveys include additional questions to aid in distinguishing formal from informal suppliers, such as the size of the business, the products or services sold, the nature of the location(s) from which sales are transacted (e.g., a physical store, an outside market, or door to door), and bookkeeping practices. The overall weighted net earnings reported by those respondents who have been classified as informal suppliers then serves as the estimate of aggregate informal supplier income. As with the earlier IRS methodology, such an approach requires specialized and potentially costly surveys, and it relies on a rather nebulous distinction between formal and informal transactions.

In contrast to these approaches, our methodology is based on publicly available information, and it does not require a distinction between formal and informal transactions. The next section provides an overview of the main steps in our methodology. This is followed by a more detailed description of the approach and our findings in the following two sections. We next perform some exercises to validate our methodology and extend the range of industries to include the general sole proprietor population. Some concluding remarks are offered in the final section.

BASIC METHODOLOGICAL STEPS

As described below, there are five main steps in our approach.

Step 1: Selection of Industry Categories with a High Informal Supplier Concentration

The first step of our methodology is the selection of industry categories that seem likely to be dominated by informal suppliers. Based on our review of the existing literature on goods and services provided by informal suppliers, we have identified 12 broad industry categories that we believe account for the vast majority of informal supplier activities.⁶ As indicated in Table 1, our list includes: food catering and roadside stands; direct sales; building maintenance/landscaping; forestry, fishing, hunting, and trapping; arts and entertainment; construction; teaching/lessons; care of children and elderly; personal services; auto repair and maintenance; other repair and maintenance; and transportation and moving.

Step 2: Development of Raw CPS-Based Estimates of Earnings

For 11 of our 12 selected industry categories, we are able to develop suitable estimates of net self-employment earnings using publicly available information from the Current Population Survey (CPS). We begin by identifying the detailed CPS industry codes that compose each broad industry category. Next, we select those survey respondents who report self-employment income associated with any of these industry codes (either through a primary or secondary job) and aggregate their reported net earnings. This serves as our preliminary raw estimate of net overall earnings within the 11 selected industry categories.

⁵ For an excellent survey of this approach, see International Labour Organization (2013).

⁶ For insights on informal supplier industries, see McCrohan and Smith (1986), McCrohan and Sugrue (1988), and Losby et al. (2002). Based on CPS data matched to Social Security Administration Detailed Earnings Records (DER), Roemer (2002) identified the 56 CPS occupation categories with the highest estimated likelihood of “underground workers” (individuals who report earnings on the CPS but who have no record of earnings in the

Step 3: Adjustment for Self-employment Income Misreported as Wages

A known issue with the CPS is that some individuals misreport their self-employment income as wages (Roemer, 2002). To address this issue, we begin by identifying the detailed CPS occupation codes that compose each of the 11 selected industry categories. We then apply an industry category-specific misreporting rate based on the Roemer (2002) study to estimate the portion of reported wages from the occupations associated with each industry category that are actually self-employment earnings. These estimated earnings are added to our raw estimate from Step 2 to produce our final estimates of net self-employment income within each of the 11 selected industry categories with a high concentration of informal suppliers.

Step 4: Use of CES Data for Food Catering and Roadside Stands

We are not able to use the CPS to estimate self-employment earnings for one industry category, food catering and roadside stands, because the industry and occupation codes in the CPS for this category are unacceptably broad. Instead, we estimate the gross receipts of vendors within this category based on tabulations of relevant consumer purchases from the Consumer Expenditure Survey (CES). These results are discussed separately in a later section.

Step 5: Estimation of Reporting Gap in Industries with High Informal Supplier Concentration

To estimate the extent of income tax evasion within industries with a high concentration of informal suppliers, we compare our estimates of actual sole proprietor earnings from our selected industry categories with the corresponding amounts reported on tax returns. For this study, we have been fortunate to have access the tax year 2001 NRP data base to estimate overall reported self-employment earnings within the relevant industry categories. In practice, however, it is not necessary to employ a sample of audited tax returns for this purpose, because

DER). The great majority of these occupations are concentrated within our 12 selected industry categories, which serves as a nice validation check on our selected industries.

comparable estimates of reported income can be obtained from a sample of unaudited returns so long as one has a reasonable indicator of the industries in which self-employed taxpayers are operating (such as the NAICS code reported by Schedule C filers in the U.S.). An advantage of using the tax year 2001 NRP audit sample is that we are able to compare the estimates of income underreporting based on our methodology to audit-based estimates. Although we work with tax year 2001 data, our methodology is of course applicable to more recent data as they become available.⁷

In order to derive estimates of reported self-employment earnings within our selected industry categories using the NRP, we have developed a “crosswalk” between the relevant CPS and NRP industry codes that compose each category.⁸ Using this crosswalk, we are able to tabulate both the overall amount of sole proprietorship earnings reported within the relevant NRP industry codes and the amount that the examiner determined should have been reported. A comparison of our CPS-based estimate of the amount of income actually earned with the NRP-based estimate of the amount of income reported yields our estimate of the income reporting gap within each informal supplier industry category. A comparison of this measure against the NRP audit-based measure of misreporting allows us to validate our reporting gap estimate.

An example of our crosswalks is presented in the Appendix for the largest industry category: #6 (Construction).⁹ All crosswalks are available upon request.

⁷ It should be noted that our analysis is restricted to informal vendors who receive money income for the goods and services they provide. We do not address the question of how to estimate noncompliance relating to barter income, which has been included in past IRS estimates of the informal supplier tax gap. For instance, estimated barter income represented \$7.3 billion of the estimated \$62.1 billion in net informal supplier income in tax year 1988 (Internal Revenue Service, 1996a).

⁸ We have validated our industry crosswalk against crosswalks developed by the U.S. Census Bureau between the 1990 Census industry codes used in the CPS and 1997 NAICS industry codes used for the NRP. See http://www.cdc.gov/niosh/topics/coding/docs/Ind_Census_1990_2000_2002_2007.xls.

⁹ The NRP crosswalks only list industry codes that are actually present in the NRP data.

CPS ESTIMATES OF INCOME FOR 11 INDUSTRY CATEGORIES

We employ data from the Current Population Survey (CPS) to estimate net sole proprietor earnings within 11 of our 12 broad industry categories containing a high concentration of informal suppliers. Note again that we do not use CPS data for the food catering and roadside stands category because we believe that the industry and occupation codes associated with this particular category are unacceptably broad for the purposes of our analysis. For this reason, we rely on the Consumer Expenditure Survey (CES) to estimate the earnings of food catering and roadside vendors. Details of this procedure are discussed in the next section.

CPS Annual Demographic File

Our primary CPS data source is the March 2002 Annual Demographic File (ADF).¹⁰ This file contains detailed micro-level demographic, employment, and income information for some 217,000 individuals belonging to a stratified random sample of approximately 78,000 households from across the U.S. The file includes codes describing the industry and occupation of the individual's current job (as of March 2002), as well as of his or her longest job in 2001. It also contains a detailed breakdown of annual 2001 earnings by source (wages and salaries, nonfarm self-employment, or farm self-employment). Separate earnings figures are provided for the individual's longest job and for all other jobs he or she held in 2001.¹¹ A code on the file identifies whether the individual was an unincorporated sole proprietor. In addition, researchers at the U.S. Census Bureau have imputed tax year 2001 federal filing status and other tax information onto the file using the comprehensive income and demographic information contained in the survey. Sample weights are available to make statistics computed from the

¹⁰ More recent versions of this source are now referred to as the Annual Social and Economic Supplement (ASEC).

¹¹ The earnings information in the CPS is top-coded for individuals who have high levels of income. We assume that the earnings of most informal suppliers fall below the relevant threshold, so that we make no adjustment for top-coding in our analysis.

survey representative of the general U.S. population in 2001 so that they can be compared with figures from the NRP, which also covers the 2001 period.¹²

Supplemental CPS Data Sources

Using the ADF, it is possible to identify individuals with self-employment earnings from a longest job in one of our selected industry categories. What is more challenging is to identify individuals who held a second job in one of these categories. A respondent may have held a second job either because he or she changed jobs during the year, or because he or she “moonlighted” (held down more than one job at the same time).

To identify job changers, we compare the industry code for the job reported at the time of the interview (March 2002) to the code for the longest job held in 2001. To identify moonlighters, we rely on supplemental information from relevant monthly CPS surveys. Of particular interest, the March 2002 CPS file contains supplementary information for each member of a large subsample of the March 2002 ADF, specifically for 156,821 individuals from 55,498 households.¹³ Further, when properly weighted, this subsample of the ADF is representative of the overall U.S. population, just like the entire ADF.

Individuals in the monthly CPS files are sampled for four consecutive months before rotating out of the sample. For the outgoing rotation group, which constitutes roughly 25 percent of the overall sample, the monthly file identifies the industry and occupation codes not only for the individual’s main job, but also for his or her second job (if any). The coding system is the same as that used for the ADF. Therefore, the March file contains the desired supplementary

¹² More specifically, the ADF universe is the civilian non-institutional population of the United States living in housing units and members of the Armed Forces living in civilian housing units on a military base or in a household not on a military base.

¹³ In addition to the March monthly CPS sample of households, the 2001 ADF contains supplemental samples to improve the accuracy of statistics on Hispanics as well as state-level estimates of children’s health insurance coverage. By applying the appropriate sample weights, statistics from either the March monthly sample or the full ADF can be made representative of the overall U.S. population.

information about an individual's second job (if any) for roughly one-fourth of the 156,821 individuals in the representative ADF subsample. In principle, supplementary information for another fourth of the ADF subsample should be available in each the monthly CPS files from April to June, so that details on the second job (if any) can be obtained for essentially all members of the ADF subsample.

In practice, however, we are only able to match information for 130,558 individuals, or 83 percent of the ADF subsample. This is largely due to sample attrition, whereby certain individuals dropped out of the sample prematurely (e.g., they changed their residence). As well, the monthly CPS files do not contain a unique code that can be used to definitively link individuals to their records in the ADF. Rather, a set of variables common to the monthly files and the ADF are used to match individual records.¹⁴ Although our matching procedure works well, it is not perfect, so that some individuals who are present on both the monthly file and the ADF may not be successfully matched. It is therefore necessary to adjust the sample weights to make our matched ADF subsample of 130,558 individuals broadly representative of the overall U.S. population.¹⁵

Result: CPS Measure of Reported Self-Employment Income

We begin by summarizing our CPS-based methodology for developing a raw estimate of aggregate net 2001 self-employment income among unincorporated sole proprietors in 11 of our broad industry categories. We then introduce a refinement to account for self-employment

¹⁴ The variables used for matching include the household identification number, the person line number, gender, and age. When matching the March monthly sample to the ADF, we also compare the recorded values of the current industry codes.

¹⁵ It is possible to match essentially all of the outgoing rotation groups from the March monthly sample to the ADF. For each subsequent rotation group from April to June, the group sample weights are proportionally adjusted upwards to account for members of the group that are not successfully matched, either because of attrition or imperfections in the matching criteria. There is a small discrepancy in the aggregate weighted populations between the ADF and the March monthly file (282.1 million compared to 278.1 million). Therefore, a small final

earnings that were erroneously reported on the CPS as wages. Our analysis is restricted to individuals who, on the basis of their reported information, either appear to have a filing requirement or are otherwise likely to file.¹⁶

Using the ADF, we are able to identify cases where an individual reports self-employment earnings in 2001 from a longest job that falls into one of our 11 selected industry categories. We are also able to determine whether an individual reports self-employment income from a second job in 2001. Unfortunately, however, no details are available about the industry or occupation associated with the second job, so we cannot determine whether it belongs to one of our selected industry categories. As discussed above, we instead rely on industry codes for second jobs that were held at a somewhat later date (e.g., at the time of an interview conducted between March and June 2002).

To estimate 2001 net self-employment income in the case of a longest job, we rely directly on the earnings reported in the ADF. However, in the case of a second job it is necessary to impute earnings. To impute self-employment earnings from a second job, we begin by computing the ratio of typical hours worked at the second job to typical hours worked at the main and second jobs combined based on the information collected from the relevant monthly CPS file. We then apply this ratio to earnings from the main job in 2001 to arrive at a preliminary estimate of earnings from the second job in 2001. Next we develop an adjustment factor to account for differences in compensation rates for the two jobs by comparing total self-employment earnings from a second job in 2001 to total earnings from a main job in 2001

proportional adjustment is applied to all matched monthly records (multiplication by 1.014) to make the weighted population total equal to the corresponding ADF total.

¹⁶ Such households are identified on the basis of the imputed variable *filestat* in the ADF, which identifies potential filing units along with their likely filing status (single, head of household, or married joint). Consistent with federal filing requirements, households that report more than \$400 in net self-employment income are flagged by this imputed variable as potential filing units. See O'Hara (2005) for a discussion of the CPS methodology for imputing potential tax filing units.

(among those reporting self-employment earnings from a main job in the ADF).¹⁷ This factor is then applied to our preliminary estimate of earnings from the second job in 2001 to produce our final estimate.

The results of our analysis are summarized in Table 2. Aggregate reported net self-employment earnings from a longest job in our 11 selected industry categories are \$105.8 billion. Estimated self-employment earnings from a second job (for job changers and moonlighters combined) contribute an additional \$8.1 billion, for a total of \$113.9 billion.

Adjustment for Self-employment Income Misreported as Wages

Roemer (2002) presents evidence that some individuals in the CPS misreport their self-employment income as wages. In particular, he identifies a substantial number of cases where a CPS respondent reports earning wages from his longest job but where matched administrative records from the Social Security Administration show only self-employment income for the respondent. Further, he finds that the problem is especially pronounced among many of the occupations within our selected industry categories.

Accordingly, we attempt to account for misclassified self-employment income within our selected industry categories, restricting our attention to households in the CPS that either appear to have a legal filing obligation or who are otherwise likely to file (such as those who appear eligible for the Earned Income Credit). In particular, we compute an explicit estimate for all potential filing units of the net amount of self-employment income in our industry categories that has been misreported in the CPS as wages.

We begin by developing a CPS-based estimate of the total reported wages within each of our 11 industry categories that are attributable to the following sources: the longest jobs held by

¹⁷ Our adjustment factor equals 0.69, indicating that net hourly earnings from the second job are on average about 69 percent as high as net hourly earnings from the main job.

primary and secondary filers, and second jobs held by primary and secondary filers (both job changers and moonlighters).¹⁸ We do so by identifying wage-earners who are associated with the CPS occupation codes defined in our crosswalk. Based on the results of Roemer (2002), we have developed an estimate of the “misreporting rate” for each of our 11 industry categories, defined as the percentage of wage and salary reports that represent misreported net self-employment earnings. These estimates are obtained by linking each of our industry categories to the relevant occupations listed in Roemer (2002) (Table 10, p. 27). Our estimate of the misreporting rate for an industry category is then computed as the ratio of wage reports involving misreported self-employment earnings to total wage reports within the linked occupation categories.

For example, we are able to link the following occupations from Roemer (2002) to our personal services industry category: barbers; child-care workers, private; and hairdressers and cosmetologists. Overall, 361,000 out of the roughly 3 million weighted respondents who reported wages from these occupations were found to have misreported their net self-employment earnings as wages, implying a misreporting rate of approximately 12 percent.¹⁹ For each industry category, we then apply the estimated misreporting rate to wages reported in the relevant occupation categories to arrive at our estimate of net self-employment earnings that are misreported as wages. The results are summarized in Table 2. Overall, we estimate that \$45.6 billion in net self-employment income is misreported as wages in the CPS.

Applying this result, we derive an adjusted CPS-based estimate of aggregate net self-employment income for our 11 industry categories (Table 2) of \$159.5 billion. A comparison of

¹⁸ Our imputation of earnings from a second job among those with misclassified income follows a process comparable to that described earlier for respondents who reported self-employment earnings from a second job.

¹⁹ For two of our categories, we are unable to link the occupations listed in Table 10 of Roemer (2002) and our industry codes. In the case of the “other repair and maintenance” category, we apply the estimated misclassification rate estimated for the “auto repair and maintenance” category, as we believe that this industry is comparable. In the case of the “teaching/lessons” industry category, we apply the estimated misreporting rate for the overall population (1.69 percent) because we consider teaching a fairly typical industry within the population.

this figure to the corresponding amount of net self-employment earnings actually reported on tax returns would yield our proposed measure of noncompliance.

Comparison with NRP Estimates of Net Self-employment Income

Although all existing methods for estimating tax noncompliance have significant limitations, IRS estimates based on the NRP are widely considered the “gold standard”. Indeed, the Bureau of Economic Analysis relies extensively on those estimates to adjust the National Accounts for the underreporting of income on tax forms, including a very sizeable adjustment to account for unreported nonfarm proprietor income.²⁰

How do our estimates compare with current IRS estimates based on the NRP? The NRP data allow us to identify members of 11 selected industry categories using either the industry codes originally reported on the tax return or the industry codes as assessed by the examiner during the audit. It seems plausible that self-employed taxpayers would tend to identify the nature of their industry comparably on their tax returns and on the CPS survey. If so, a comparison of reported earnings on the two data sources would provide a tentative estimate of reporting noncompliance among individuals who perceive themselves to be operating in industries dominated by informal suppliers.²¹

This comparison is provided in Table 3. Overall, the results indicate that taxpayers reported approximately \$55.7 billion in aggregate net self-employment earnings within the 11 selected industry categories on their tax year 2001 federal individual income tax returns; after accounting for additional self-employment income that was uncovered during the NRP examinations (or the audit-adjusted amount), overall estimated earnings amount to \$98.4

²⁰ For more details on these adjustments, refer to Congressional Budget Office (2013) and Bureau of Economic Analysis (2013).

²¹ Of course, some taxpayers may strategically misstate their industry code on the tax return if they perceive that the actual industry code may draw a higher level of scrutiny from the tax authority.

billion.²² Even this audit-adjusted figure is well below our CPS-based estimate of \$159.5 billion. Furthermore, the CPS-based estimate is consistently higher across all industry categories, suggesting that our survey-based approach is able to uncover more income underreporting than even an intensive random audit program.

Overall, the largest overall dollar gap between our CPS-based estimates and the amount reported on federal individual income tax returns is associated with the construction industry category (\$30.2 billion), followed by care of children and the elderly (\$15.8 billion), and transportation and moving (\$10.5 billion). When looking instead at the ratio of the tax return estimates to our CPS-based estimates, the industry categories with the lowest percentage of income reported on tax returns are forestry, hunting, fishing, and trapping (13.7 percent), teaching/lessons (17.3 percent), and arts and entertainment (17.3 percent).

An additional IRS estimate is also of interest. The IRS attempts to account for the difficulties NRP examiners experience in uncovering underreported income on tax returns using an econometric methodology known as “Detection Controlled Estimation” (DCE).²³ Applying the DCE estimate of the detection error rate for self-employment earnings to NRP data yields a detection-adjusted estimate of net self-employment earnings within the 11 selected industry categories of \$209.1 billion. Now our CPS-based estimate of \$159.5 billion in net self-

²² The audit-adjusted amount includes adjustments for cases in which self-employment income is reported on the return but on the wrong line item (e.g., when self-employment earnings are improperly reported as wages).

²³ DCE is a statistical methodology developed by Feinstein (1990) to account for imperfections in examination processes (e.g., audits) to fully uncover violations (e.g., tax evasion). Under this methodology, one jointly models the detection process along with the underlying violation of interest. See Erard and Feinstein (2012) for a discussion of the application of the DCE methodology to the tax gap, in which a “multiplier” is applied to the initial estimate of noncompliance to generate the DCE estimate. Note that negative examiner adjustments to reported income amounts (or apparent overstatements) are not modified by the multiplier.

employment income is about 24 percent lower than the detection-adjusted estimate of \$209.1 billion.²⁴

Our CPS results also appear to understate the overall number of informal suppliers to some extent. In particular, our CPS-based statistics indicate that about 7.36 million households with non-zero self-employment earnings among our 11 industry categories should have filed a tax year 2001 federal individual income tax return, which is somewhat larger than our NRP-based estimate of the number of returns with non-zero reported self-employment earnings that were actually filed (7.04 million). We would have expected this discrepancy to be even larger, since some sole proprietors fail to comply with their filing obligation and others fail to report their self-employment earnings when they do file.²⁵ Overall, the modest size of this discrepancy may be an indication that some relatively low-earning informal suppliers, such as undocumented immigrants, are reluctant to report their income on the CPS survey. Still, our CPS-based estimate of the magnitude of net self-employment earnings of \$155.9 billion exceeds the \$98.4 billion in net earnings that NRP examiners are able to identify through reasonably intensive audits (while falling short of the DCE estimate of \$209.1 billion). Thus, our methodology appears to be a useful tool for using publicly available information to derive a meaningful lower bound for market activities in industries dominated by informal suppliers and the extent of income tax evasion associated with these activities.

²⁴ In our view, the DCE methodology serves as a credible way to overcome the detection problem associated with unreported self-employment income when one has access to a comprehensive random sample of audited tax returns, such as the NRP. In this regard, the portion of the DCE-adjusted underreporting gap associated with industries with a high concentration of informal suppliers serves as a meaningful indicator of unreported informal sector activity and avoids the need to make a rather dubious distinction between formal and informal suppliers. In cases where comprehensive random audit data are not available, our methodology produces a somewhat more conservative estimate using only public survey information and basic tax return statistics.

²⁵ Some Schedule C filers report zero net earnings within these industry categories; based on the examiner assessments, about one-half million such filers should have reported non-zero income.

After accounting for undetected noncompliance, the NRP estimates indicate that the 11 selected industry categories account for about 39 percent of total Schedule C net self-employment income (with the remaining 61 percent attributable to earnings among industries dominated by formal suppliers), but for about 46 percent of total Schedule C net income underreporting. These results are consistent with the notion that informal supplier industries account for a disproportionate share of noncompliance. They are also consistent with the notion that the industries we selected are responsible for a disproportionate share of all noncompliance.

Our estimation methodology relies on self-employed workers providing reasonably accurate responses on the CPS survey and on their tax returns regarding the industries in which they operate so that we can properly identify those participating in our 11 selected industry categories. Using the tax year 2001 NRP results, we are able to investigate the degree to which taxpayer self-reports of industry categories on federal individual income tax returns are consistent with the assessments of NRP examiners. The weighted number of Schedule C forms reporting industry codes pertaining to the 11 selected industry categories is 8.36 million.²⁶ Among these returns, the NRP examiners agree with this classification in 7.71 million cases (or about 92 percent of the time), meaning that the vast majority of taxpayers who self-identify themselves as operating in industries with a high concentration of informal suppliers are actually operating in these industries. On the other hand, the NRP examiners have identified an additional 1.88 million sole proprietorships from one of the 11 industry categories that either have reported an industry code outside of these categories or have failed to report self-employment earnings altogether, implying an overall sub-population of 10.2 million sole proprietorships. Thus, some

²⁶ Note that this figure exceeds the number of taxpayers reporting non-zero self-employment earnings within these industry categories, both because some taxpayers file multiple Schedule C returns and because some Schedule C returns report zero net income.

informal suppliers will tend to go uncounted under our methodology, making our overall estimate of activity within the relevant industries a conservative one.

CES ESTIMATES OF EARNINGS FROM FOOD CATERING AND ROADSIDE STANDS

CES Estimates

As noted earlier, we are not able to use the CPS to estimate self-employment earnings for one industry category, food catering and roadside stands, because the industry and occupation codes in the CPS for this category are unacceptably broad.²⁷ Instead, we estimate the gross receipts of vendors within this category based on tabulations from the Consumer Expenditure Survey (CES).²⁸

The CES is conducted by the Bureau of Labor Statistics in the U.S. Department of Labor, and it provides detailed information on the expenditure patterns of American consumers (information that is also used to revise the Consumer Price Index). The survey consists of two separate components: a quarterly “Interview Survey” in which each consumer unit in the sample is interviewed every three months over a 15-month period, and a “Diary Survey” completed by subsample of consumer units for two consecutive one-week periods. We rely on the Interview Survey for our analysis. It has the advantage of including a much larger sample of respondents who report purchases from catering and roadside vendors. The Bureau of Labor Statistics estimates that 90 to 95 percent of total consumer expenditures are covered in this survey.

Our estimate of consumer expenditures on food catering and roadside stands is based on the Detailed Expenditure Files (DEF) from the 2001 CES Interview Survey. Included under

²⁷ Food catering is subsumed under the CPS restaurants and other food services industry category, while roadside stands fall within the CPS crop production category.

²⁸ Note that we use here the “Schedule C Principal Business or Professional Activity (NAICS) Codes” of:

“Miscellaneous Expenses” in the DEF are expenditures made on “Catered Affairs”. Similarly, under “Expense Patterns for Food, Beverages, and Other Selected Items” are consumer expenditures on vegetable stands and farmers’ markets.²⁹ On an annual basis, 2001 total expenditures are \$4.13 billion for food catering and \$1.68 billion for roadside stands, for a combined total of \$5.81 billion. See Table 4.

Comparisons with Other Estimates

It is worth noting that our estimates of expenditures on catering and roadside stands are comparable to those from other, independent sources. For example, the National Restaurant Association (2004) estimated that 2001 expenditures on social and mobile caterers totaled \$4.8 billion. Further, the U.S. Department of Agriculture (2004) estimated that the “value of agricultural products sold directly to individuals for human consumption” was \$812 million in 2002.

Our estimates are also very similar to current estimates of gross income derived from the NRP. The NRP per exam measure of gross income, which accounts for any audit adjustments to amounts originally reported by the taxpayer, is \$5.84 billion; the net income per exam estimate is \$734.7 million; and the detection-adjusted NRP net income amount is \$1.92 billion (obtained by applying the estimated NRP detection-based multiplier for Schedule C income only to positive examiner adjustments to reported net income amounts). If we assume that all of the detection adjustment to the raw NRP per exam net income amount is attributable to undiscovered gross income (i.e., that overreporting of expenses was perfectly detected since the burden of proof was

722300 (Special food services including food service contractors & caterers) and 445230 (Fruit and vegetable markets). Also, we use the “NRP Occupation Code” of 521 (Food & beverage preparation/service occupation).

²⁹ The exact description of the expenditure variable is “Quarterly expenditure for food or nonalcoholic beverages from places other than grocery stores, such as home delivery, specialty stores, bakeries, convenience stores, dairy stores, vegetable stands, or farmers’ markets”.

on the taxpayer for these items), we have that the detection-adjusted gross income for food catering and roadside stands equals [\$5.84 billion + 1.92 billion – 0.73 billion], or \$7.03 billion.

As a result, our CES-based measure of gross income among food catering and roadside stands is almost identical to the raw NRP per exam measure (\$5.84 billion), and it is about 17 percent less than the detection-adjusted NRP measure (\$7.03 billion). Our measure is meant to pick up underreporting by both filers and nonfilers, whereas the NRP figures only apply to filers. As with the other 11 industry categories, our measure probably understates overall noncompliance by filers and nonfilers. Even so, it performs fairly well.

VALIDATION AND EXTENSION OF METHODOLOGY

As discussed earlier, our methodology relies on reasonably good reporting of self-employment earnings by survey respondents. So far, we have established that the survey responses within our 11 selected industry categories reveal substantially more self-employment income than is indicated on tax returns, even after accounting for adjustments made by examiners during intensive random audits. At the same time, our estimates fall somewhat short of IRS estimates of underreporting that attempt to account for income that goes undetected during tax audits. This suggests that our estimate of underreporting may best be viewed as a lower bound on noncompliance within the selected industry categories. Such a view is supported by the findings of Pissarides and Weber (1989), who found evidence of significant underreporting of self-employment earnings within the Family Expenditure Survey in the U.K. In order to more directly examine the degree to which CPS respondents are forthcoming about their earnings in the U.S., we undertake some additional validation exercises.

Comparison of CPS and Tax Return Reports

Our first validation exercise is to compare reporting of wage and salary earnings in the CPS with wage and salary reporting on income tax returns. Tax compliance with respect to the reporting of this income source is believed to be extremely high, owing to the presence of third-party reporting and tax withholding. Therefore, a comparison over time of wage and salary reporting on the CPS with reporting on income tax returns provides a check on the accuracy of CPS respondents in reporting the other major source of earned income besides self-employment earnings.³⁰ Table 5 provides such a comparison for tax years 1996 through 2012. The CPS estimates include all reported wage, salary, and tip earnings, including any amounts earned while moonlighting. The results show that reported earnings are consistently somewhat higher in the CPS. This makes sense, both because the CPS population includes both filers and nonfilers of federal individual income tax returns³¹ and because some filers underreport their wages, salaries, and, especially, tips, to some degree. Overall, then, it appears that these earnings are accurately reported by CPS respondents, which demonstrates that they are capable and willing to be forthcoming, at least with respect to reporting earnings from this source.

Comparison of CPS and Tax Return Reports for All Industries

Our methodology assumes that CPS respondents are also reasonably forthcoming with regard to their reporting of self-employment earnings in the 11 selected industry categories. If so, it should also be the case that they are reasonably forthcoming with regard to reporting self-employment earnings in other industries. To investigate this issue, we compare in Table 6 CPS reports of overall net sole proprietor earnings across all industries with reports on federal individual income tax returns.³² The CPS estimates include reported self-employment income

³⁰ More precisely, we use the Current Population Survey Annual Social and Economic Supplement (CPS ASEC).

³¹ Note that Cilke (2014) estimates that nonfilers earned wages and salaries of \$108 billion in tax year 2010.

³² Again, we use the Current Population Survey Annual Social and Economic Supplement (CPS ASEC).

from all work, including moonlighting.³³ The results indicate that CPS estimates are consistently much larger than the amounts reported on tax returns, consistent with our earlier findings for the 11 selected industry categories. Over the years, the CPS estimate of overall self-employment earnings ranges from 19 to 54 percent higher than amount reported on federal individual income tax returns, which is suggestive of a considerable degree of tax noncompliance. While some respondents, such as undocumented immigrants, may be reluctant to report their earnings even on a confidential survey, CPS respondents overall seem to be fairly responsive.

Comparison of CPS and Social Security Earnings Records

Roemer (2002) carries out an exact match between earnings reported by respondents in the CPS and their Detailed Earnings Records (DER) from the Social Security Administration. His results show that many CPS respondents, particularly those employed in informal occupations, report earnings on the CPS that are not recorded in the DER, suggesting that the CPS does indeed pick up a fair amount of underground earnings.

SUMMARY AND CONCLUSIONS

Our proposed methodology for estimating the earnings of informal suppliers has a number of advantages over existing approaches, including:

- In contrast to many of the approaches used to indirectly measure the underground economy, our approach relies on a direct comparison of income earned and income reported on tax returns, and it is specifically targeted to industries dominated by informal suppliers;
- The survey information used in our approach is publicly available, meaning that no special surveys need to be commissioned;
- The number of respondents to the surveys used in our approach is much larger than the number of respondents to the special surveys used in the earlier IRS methodology;
- The difficult issue of distinguishing informal suppliers from formal suppliers is avoided;
- With the exception of the food catering and roadside stands industry category, sales to both consumers and businesses are fully accounted for;

³³ No adjustment has been made to account for respondents who misreported self-employment earnings as wages.

- With the exception of the food catering and roadside stands industry category, the approach provides a direct estimate of net earnings, thereby avoiding the need to rely on assumptions about the relationship between net income and gross receipts; and
- Detailed crosswalks have been developed that provide a tight linkage between the coding used in the surveys for selected industries and occupations and the coding used in the NRP.

The use of a direct income measure based on publicly available information is an especially compelling advantage. Also, although we have focused our analysis on the particularly vexing issue of measuring income and tax compliance in industries dominated by informal suppliers, we have also illustrated how our approach can be extended to evaluate income and compliance trends with regard to self-employment earnings across all industries. While our results do not answer all of the important questions regarding informal supplier evasion, we believe that they do provide a meaningful estimate of the scope of evasion within the key industry sectors associated with informal supplier activity, information that is of substantial importance to researchers, policy-makers, and tax administrators.

A potential drawback of our methodology is that it relies on the accuracy of income information reported in the CPS by proprietors who operate businesses in our selected industry categories. As a group, these individuals may be more willing to provide an accurate accounting of their income on an independently administered and confidential survey than they would on their tax returns. Nevertheless, it is entirely possible that the amounts reported on the CPS fall somewhat short of true earnings, in which case our methodology would tend to underestimate the extent of noncompliance.

However, the evidence on this issue is encouraging. We have demonstrated that our methodology uncovers substantially more unreported self-employment earnings in our 11 selected industry categories than IRS examiners are able to detect through intensive random audits (although somewhat less than what IRS estimates the true level of underreporting to be

when undetected noncompliance is taken into account). We also have shown that CPS respondents are very forthcoming with respect to the other key source of earned income (wages and salaries) and that CPS estimates of overall sole proprietor earnings across all industries are in fact consistently much higher than those reported on tax returns. Further, Roemer (2002) provides evidence that the CPS does in fact pick up a significant amount of underground income. Thus, while some CPS respondents, such as undocumented immigrants, may be reluctant to fully report their earnings, the available evidence suggests that most CPS respondents are reasonably forthcoming. Overall, we feel that our methodology produces a plausible, indeed somewhat conservative, estimate both of the earnings and of the degree of tax evasion within those industries dominated by informal suppliers.

To arrive at our estimate of noncompliance, we compare our survey-based estimate of self-employment earnings within industries dominated by informal suppliers to the amount reported on federal individual income tax returns for those industries. In our study we are able to access the NRP to estimate overall reported self-employment earnings within the relevant industry categories. Importantly, however, it is not actually necessary to employ a sample of *audited* tax returns for this purpose because comparable estimates of reported income can be obtained from a sample of *unaudited* returns as long as one has a reasonable indicator of the industries in which self-employed taxpayers are operating (such as the NAICS code reported by Schedule C filers in the U.S.) An advantage of using the NRP audit sample is that we are able to compare the estimates based on our methodology to audit-based estimates of underreporting.

In sum, our methodology provides a new way to estimate informal supplier income that uses information that is likely to be readily and cheaply available in some form in many countries. We believe that our approach is useful both for developed countries like the U.S. and,

subject to data quality issues, for those developing countries in which informal suppliers are likely to account for a large share of economic activity (and a large amount of tax evasion).³⁴

REFERENCES

- Alm, James. 2012. "Measuring, Explaining, and Controlling Tax Evasion: Lessons from Theory, Field Studies, and Experiments." *International Tax and Public Finance* 19 (1): 54-77.
- Blades, Derek. 2011. "Estimating Value Added of Illegal Production in the Western Balkans." *Review of Income and Wealth* 57 (1): 183-195.
- Brown, Allison. 2001. "Counting Farmers Markets." *The Geographical Review* 51 (4): 655-674.
- Brown, Allison. 2002. "Farmers' Market Research 1940-2000: An Inventory and Review." *American Journal of Alternative Agriculture* 17 (4): 167-176.
- Bureau of Economic Analysis. 2013. "National Income and Product Accounts, Gross Domestic Product: Second Quarter 2013 (Advance Estimate), Comprehensive Revision: 1929 Through First Quarter 2013." News Release. BEA 13-34. Washington, D.C. (http://www.bea.gov/newsreleases/national/gdp/2013/pdf/gdp2q13_adv.pdf).
- Cilke, James. 2014. "The Case of Missing Strangers: What We Know and Don't Know About Nonfilers." Paper presented at the 107th Annual Conference of the National Tax Association, Santa Fe, New Mexico, November 13-15, 2014.
- Congressional Budget Office. 2013. "How CBO Projects Income. Washington, D.C. (http://www.cbo.gov/sites/default/files/cbofiles/attachments/IncomeProjection_onecolumn.pdf).
- Ekici, Tufan and Mustafa Besim. 2014. "A Measure of the Shadow Economy in a Small Economy: Evidence from Household-level Expenditure Patterns." *Review of Income and Wealth* forthcoming.
- Engstrom, Per and Bertil Holmond. 2009. "Tax Evasion and Self-employment in a High-tax Country: Evidence from Sweden." *Applied Economics* 41 (19): 2419-2430.
- Erard, Brian and Jonathan S. Feinstein. 2012. "The Individual Income Reporting Gap: What We See and What We Don't." *2011 IRS Research Bulletin: Recent Research on Tax Administration and Tax Compliance*. Washington, D.C., 129-142.
- Erard, Brian and Chih-Chin Ho. 2003. "Explaining the U.S. Income Tax Compliance Continuum." *eJournal of Tax Research* 1 (2): 93-109.
- Feinstein, Jonathan S. 1990. "Detection Controlled Estimation." *Journal of Law and Economics* 33 (1): 233-276.
- Feldman, Naomi and Joel Slemrod. 2007. "Estimating Tax Compliance with Evidence from Unaudited Tax Returns." *The Economic Journal* 117 (518): 327-352.
- Gibson, John, Bonggeon Kim, and Chul Chung. 2009. "Using Panel Data to Exactly Estimate Income Underreporting of the Self-employed." Korea Institute for International Economic Policy Working Paper. Seoul, Korea.

³⁴ Some constructive guidance on the formulation and administration of surveys that would be suitable for applying our methodology in developing countries (perhaps with some refinements) can be found in Organisation for Economic Co-operation and Development (2002). See Blades (2011) for an application of some of these methods to illegal activities.

- Hurst, Erik, Geng Li, and Benjamin Pugsley. 2014. "Are Household Surveys Like Tax Forms? Evidence from Income Underreporting of the Self-employed." *The Review of Economics and Statistics* 96 (1): 19-33.
- Internal Revenue Service. 1988. "Income Tax Compliance Research, Supporting Appendices to Publication 7285." Internal Revenue Service Publication 1415. Washington, D.C.
- Internal Revenue Service. 1996a. "Federal Tax Compliance Research: Individual Income Tax Gap Estimates for 1985, 1988, and 1992." Internal Revenue Service Publication 1415. Washington, D.C.
- Internal Revenue Service. 1996b. "Tax Administration: Alternative Strategies to Obtain Compliance Data". Report to the Commissioner, Internal Revenue Service. Washington, D.C.: United States General Accounting Office.
- International Labour Organization. 2013. "Measuring Informality: A Statistical Manual on the Informal Sector and Informal Sector Employment." Geneva: International Labour Organization.
- Johannson, Edvard. 2005. "An Estimate of Self Employment Income Underreporting in Finland." *Nordic Journal of Political Economy* 31 (1): 99-109.
- Joshi, Kaushal, Glenita Amoranto, and Rana Hasan. 2011. "Informal Sector Enterprises: Some Measurement Issues." *The Review of Income and Wealth* 57 (1): S143-S165.
- Kulshreshtha, A. C. 2011. "Measuring the Unorganized Sector in India." *The Review of Income and Wealth* 57 (1): S123-S134.
- Losby, Jan L., John F. Else, Marcia E. Kingslow, Elaine L. Edgcomb, Erika T. Malm, and Vivian Kao. 2002. "Informal Economy Literature Review." Institute for Social and Economic Development and the Aspen Institute (<http://www.fieldus.org>).
- Luttikhuisen, Ronald and Brug Kazemier. 2000. "A Systematic Approach to the Hidden and Informal Activities." Statistics Netherlands.
- McCrohan, Kevin F. and James D. Smith. 1986. "A Consumer Expenditure Approach to Estimating the Size of the Underground Economy." *Journal of Marketing* 50 (2): 22-40.
- McCrohan, Kevin F., James D. Smith, and Terry K. Adams. 1991. "The Cyclical Nature of Household Purchases in Informal Markets." *European Journal of Marketing* 25 (7): 22-40.
- McCrohan, Kevin F. and Timothy F. Sugrue. 1988. "An Empirical Analysis of Informal Markets Participation." *Social Science Quarterly* 79 (1): 212-26.
- National Restaurant Association. 2004. "2004 Restaurant Industry Forecast, Executive Summary." Washington, D.C. (<http://www.restaurant.org>).
- O'Hara, Amy. 2005. "Evaluation of CPS Tax Simulation Using Administrative IRS Data." Paper presented at Federal Committee on Statistical Methodology Research Conference, Arlington, VA, November 14-16, 2005.
- Organisation for Economic Co-operation and Development. 2002. *Measuring the Non-observed Economy: A Handbook*. Paris, France: OECD.
- Pissarides, Christopher A. and Guglielmo Weber. 1989. "An Expenditure-based Estimate of Britain's Black Economy." *Journal of Public Economics* 39 (1): 17-32.
- 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." U.S. Census Bureau Staff Paper. Washington, D.C.
- Schneider, Friedrich and Dominik H. Enste. 2002. *The Shadow Economy – An International Survey*. Cambridge, MA: Cambridge University Press.

- Schuetze, H. 2002. "Profiles of Tax Noncompliance among the Self Employed In Canada: 1969-1992." *Canadian Public Policy* 28 (2): 219-237.
- Slemrod, Joel and Caroline Weber. 2012. "Evidence of the Invisible: Toward a Credibility Revolution in the Empirical Analysis of Tax Evasion and the Informal Economy." *International Tax and Public Finance* 19 (1): 25-53.
- Smith, James D. and Terry K. Adams. 1987. "The Measurement of Selected Income Flows in Informal Markets, 1981 and 1985-1986." Ann Arbor, MI: Economic Behavior Program, Survey Research Center, Institute for Social Research, The University of Michigan.
- Treasury Inspector General for Tax Administration. 2013. "The Internal Revenue Service Needs to Improve the Comprehensiveness, Accuracy, Reliability, and Timeliness of the Tax Gap Estimate" (Reference Number 2013-IE-R008). Washington, D.C.: U.S Department of the Treasury (<http://www.treasury.gov/tigta/iereports/2013reports/2013IER008fr.pdf>).
- United States Department of Agriculture. 2004. "2002 Census of Agriculture." Washington, D.C. (<http://www.usda.gov/nass/>).

APPENDIX:

EXAMPLE OF CROSSWALKS BETWEEN CPS AND NRP INDUSTRY AND OCCUPATION CODES

Category 6: Construction

Schedule C Principal Business or Professional Activity (NAICS) Code

233200 Residential building construction
 233200 Residential building construction
 233210 Single family housing construction
 233300 Nonresidential building construction
 235000 Special trade contractors
 235100 Plumbing, heating, & air-conditioning contractors
 235110 Plumbing, heating, & air-conditioning contractors
 235210 Painting & wall covering contractors
 235310 Electrical contractors
 235400 Masonry, drywall, insulation, & tile contractors
 235500 Carpentry & floor contractors
 235610 Roofing, siding, & sheet metal contractors
 235710 Concrete contractors
 235810 Water well drilling contractors
 235900 Other special trade contractors
 233110 Land subdivision & land development
 234100 Highway, street, bridge, & tunnel construction
 234900 Other heavy construction

TY 2001 NRP Occupation Code

502 Day worker
 631 Supervisors: construction
 641 Brickmason/stonemason/hard tile setter
 642 Carpenters and related workers
 851 Supervisor: handlers/cleaners/helpers/laborers
 864 Helper: construction trades
 871 Construction laborer

643 Electrician/power transmission installer
644 Painter/paperhanger/plasterer
645 Plumber/pipefitter/steamfitter
646 Other construction worker
851 Supervisor: handlers/cleaners/helpers/laborers
864 Helper: construction trades
871 Construction laborer

2002 CPS Industry Classification

60 Construction

2002 CPS Occupation Classification

35 Construction inspectors
553-558 Supervisors, construction occupations
563 Brickmasons and stonemasons
564 Brickmason and stonemason apprentices
565 Tile setters
566 Carpet installers
567 Carpenters
569 Carpenter apprentices
573 Drywall installers
575 Electricians
576 Electrician apprentices
577 Electrical power installers and repairers
579 Painters, construction and maintenance
583 Paperhangers
584 Plasterers
585 Plumbers, pipefitters, and steamfitters
587 Plumber, pipefitter, and steamfitter apprentices
588 Concrete and terrazzo finishers
589 Glaziers
593 Insulation workers
595 Roofers
596 Sheetmetal duct installers
597 Structural metal workers
599 Construction trades, n.e.c.
844 Operating engineers
853 Excavating and loading machine operators
855 Grader, dozer, and scraper operators
866 Helpers, construction trades
869 Construction laborers

TABLE 1
Key Industry/Occupation Categories for Informal Suppliers

1. Food catering and roadside stands
2. Direct sales
3. Building maintenance/landscaping
4. Forestry, fishing, hunting, and trapping
5. Arts and entertainment
6. Construction
7. Teaching/lessons
8. Care of children and elderly (including home health services)
9. Personal services
10. Auto repair and maintenance
11. Other repair and maintenance
12. Transportation and moving

TABLE 2
Self-employment Income for 11 Industry Categories

Source of Income	Amount (in \$ billions)
<i>Net Self-employment Income</i>	
Longest job	105.8
Second job:	
Job changers	2.9
Moonlighters	5.2
Total Net Self-employment Income	113.9
<i>Misclassified Self-employment Income</i>	
Longest job	44.1
Second job:	
Job changers	0.8
Moonlighters	0.7
Total Misclassified Self-employment Income	45.6
Total	159.5

Source: Authors' calculations.

TABLE 3
Comparison of Estimates of Self-Employment Income for 11 Industry Categories:
CPS, Tax Returns, and Audit-adjusted Estimates

Industry Category	CPS Estimate		Tax Returns Estimate*		Audit-adjusted Estimate	
	Net Income (in \$ billions)	Standard Error	Net Income (in \$ billions)	Standard Error	Net Income (in \$ billions)	Standard Error
Direct sales	4.2	1.03	0.9	0.37	2.4	0.39
Building maintenance/landscaping	10.4	0.86	3.3	0.35	6.1	0.49
Forestry, fishing, hunting, and trapping	2.4	0.37	0.3	0.22	1.2	0.24
Arts and entertainment	11.4	0.77	2.0	0.64	4.5	0.70
Construction	53.3	2.52	23.2	1.09	40.9	1.47
Teaching/lessons	7.5	0.72	1.3	0.25	1.8	0.28
Care of children and elderly	20.1	0.93	4.3	0.40	7.0	0.51
Personal services	19.4	1.46	10.7	0.67	16.0	0.79
Auto repair and maintenance	8.4	0.79	2.0	0.31	4.2	0.41
Other repair and maintenance	6.7	0.62	2.3	0.32	4.1	0.43
Transportation and moving	15.8	1.05	5.4	0.41	9.9	0.55
Total	159.5	3.81	55.7	1.45	98.4	1.76

Source: Authors' calculations.

*The tax returns estimate has been corrected by the NRP examiners to include self-employment earnings erroneously reported as income from another source, such as wages.

TABLE 4
Gross Self-employment Income for Food Catering and Roadside Stands

Source of Income	Amount (in \$ billions)
Food catering	4.13
Roadside stands	1.68
Total	5.81

Source: Authors' calculations.

TABLE 5
Overall Reported Wages on the CPS and on Federal Individual Income Tax Returns for
Tax Years 1996-2012*

Tax Year	CPS Estimate (in \$ thousands)	Tax Returns Estimate (in \$ thousands)	Ratio of CPS to Tax Returns Estimates
1996	3,660,938,603	3,376,871,545	1.08
1997	3,846,231,046	3,613,918,456	1.06
1998	4,095,034,174	3,879,762,259	1.06
1999	4,276,710,059	4,132,473,459	1.03
2000	4,701,595,980	4,456,167,438	1.06
2001	4,978,034,902	4,565,229,218	1.09
2002	5,072,982,153	4,559,690,903	1.11
2003	5,159,559,025	4,649,900,493	1.11
2004	5,350,760,544	4,921,806,344	1.09
2005	5,614,721,422	5,155,407,373	1.09
2006	5,963,713,453	5,469,370,119	1.09
2007	6,135,610,878	5,842,269,820	1.05
2008	6,236,407,846	5,950,634,829	1.05
2009	6,130,674,721	5,707,088,487	1.07
2010	6,132,916,291	5,837,350,365	1.05
2011	6,456,980,291	6,055,389,434	1.07
2012	6,682,453,979	6,301,357,591	1.06

Source: www.irs.gov/pub/irs-soi/12intaba.xls).

* The figures are not adjusted to account for self-employment earnings misreported as wages.

TABLE 6
Overall Reported Sole Proprietor Earnings on the CPS and on Federal Individual
Income Tax Returns for Tax Years 1996-2012*

Tax Year	CPS Estimate (in \$ thousands)	Tax Returns Estimate (in \$ thousands)	Ratio of CPS to Tax Returns Estimates
1996	234,914,037	176,903,956	1.33
1997	288,097,753	186,741,216	1.54
1998	292,202,925	202,400,115	1.44
1999	292,288,538	208,414,067	1.40
2000	312,071,339	213,865,353	1.46
2001	289,913,774	216,772,496	1.34
2002	308,060,458	220,783,572	1.40
2003	329,721,196	229,655,285	1.44
2004	322,108,713	247,217,287	1.30
2005	382,245,360	269,701,056	1.42
2006	411,670,475	281,527,260	1.46
2007	396,047,718	279,736,260	1.42
2008	379,138,120	264,234,283	1.43
2009	338,609,878	244,982,827	1.38
2010	344,800,644	267,265,585	1.29
2011	344,207,900	282,969,817	1.22
2012	363,209,682	304,191,539	1.19

Source: www.irs.gov/pub/irs-soi/12intaba.xls).

* The figures are not adjusted to account for self-employment earnings misreported as wages.