Using Public Information to Estimate Informal Supplier Income

James Alm and Brian Erard*

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 informal supplier income. Our methodology involves using national survey results on self-employment earnings within a carefully selected set of industry categories where informal activities are 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 those of formal suppliers within these categories. Then, by comparing these national survey results on self-employment earnings within selected industry categories to Internal Revenue Service statistics on the amounts actually reported for tax purposes, it is possible to estimate the extent of informal supplier tax noncompliance. We illustrate our approach using data from several years.

JEL classification: H26; C81.

Keywords: Tax evasion; informal suppliers.

* 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. 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).
1. Introduction

An enduring problem in the analysis of tax evasion is the difficulty of its measurement. “Tax evasion” refers to illegal and intentional actions taken by individuals to reduce their legally due tax obligations through such means as: underreporting income; overstating deductions, exemptions, or credits; failing to file appropriate tax returns; or even by engaging in barter. There is widespread belief that these many avenues for tax evasion are well traveled within most tax jurisdictions across the globe. However, this “widespread belief” is necessarily imprecise because reliable information on taxpayer compliance is difficult to obtain. After all, tax evasion is illegal, and individuals have strong incentives to conceal their cheating. There have been many approaches to measurement, often using “direct”, “indirect”, and “model” approaches; more recently, researchers have been increasingly creative in devising “modern” methods of measurement. All methods are subject to criticism and, in many cases, a good deal of skepticism.¹

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

¹ See Schneider and Enste (2000, 2002) for detailed discussions and critiques of these approaches. For a more recent discussion, see Alm (2012).
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?  

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 ones like those performed under the National Research Program (NRP) or its predecessor, the Taxpayer Compliance Measurement Program (TCMP). In the past, the IRS has attempted to address this potentially severe non-detection problem with the aid of supplementary information from a special survey it periodically commissioned of consumer purchases in the informal sector, most recently 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.

---

2 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). Useful surveys are in Losby et al. (2002) and OECD (2002). See also IRS (1996b).

3 In more recent studies, the IRS has attempted to apply statistical methods (i.e., “detection controlled estimation”) to audited returns to estimate tax evasion.
The starting point in this estimation process therefore relied upon the use of a survey of 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.

Rather than rely on a rather dubious distinction between formal and informal transactions, we develop in this paper a new approach for estimating informal supplier income, which can then be used as the first step in estimating informal supplier tax evasion. More specifically, our methodology involves using national survey results on self-employment earnings within a carefully selected set of industry categories where informal activities are 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 selected industry categories to statistics on the portion of these earnings ultimately reported for tax purposes, it is possible to estimate the overall extent of informal supplier tax noncompliance.

Importantly, our methodology for estimating informal supplier earnings does not rely on data sources that are 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 12 industry categories. Similar information is frequently available in many countries outside of the U.S., given the growing use of detailed expenditure surveys.
from organizations such as The World Bank.\textsuperscript{4} The main data requirement is tabulations of reported self-employment earnings (or consumer expenditures) within selected industry categories based on a statistically representative survey. A possible concern with our approach is the extent to which informal suppliers will accurately reveal their true earnings on a survey. The answer appears to be that they are reasonably forthcoming. In particular, we have compared our estimates against tabulations based on a large stratified random sample of intensively audited tax returns conducted under the NRP of the IRS. Our survey-based estimate of informal supplier earnings in tax year 2001 exceeds what NRP examiners were able assess based on their intensive audits; however, our estimates fall somewhat short of an NRP-based estimate that attempts to account for undetected informal supplier income.\textsuperscript{5} 

To arrive at an 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 this study we are able to access the NRP 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 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 NRP audit sample is that we are able to compare the estimates based on our methodology to audit-based estimates of underreporting.

\textsuperscript{4} For example, see the publicly available microdata survey information available at The World Bank website \url{http://microdata.worldbank.org/index.php/catalog/central}.

\textsuperscript{5} This estimate is obtained by expanding unreported income that was discovered by an NRP examiner using a multiplier estimate of the ratio of actual unreported income to detected income.
Below we lay out our methodology, present our results, and discuss the relative merits of our approach. We begin in section 2 by identifying a comprehensive set of 12 industry categories within which informal suppliers are likely to operate. We then describe the detailed “crosswalks” we have developed that link the relevant industry and occupation codes for 11 of these categories in the NRP database to the corresponding codes in our primary survey data source, the CPS. The CPS data samples that we have drawn together for our analysis of these 11 industry categories are summarized in section 3. We use these samples to present our raw estimates of self-employment earnings in these 11 selected industry categories based on the earnings reported by individuals in the CPS. We then adjust these estimates to account for self-employment income that has been misreported as wages. In section 4 we discuss an additional data source, the CES, which we use to estimate unreported income from the remaining industry category (food caterers and roadside stands) for which we believe that the CPS approach is inappropriate. We conclude in section 5 with a discussion of the relative merits of our methodology. Throughout, we apply our method to tax year 2001 data because it is for this year that NRP and other data were available. Our method is of course applicable to more recent data as they become available. Our approach is also applicable to other countries with similar data sources; specifically, survey data and tax return statistics that break down reported self-employment earnings by industry. Indeed, our approach should be especially useful for deriving estimates within developing countries in which informal suppliers are thought to be dominant.

---

6 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).
2. Selection of Industry and Occupation Categories and “Crosswalks”

The starting point of our methodology is the selection of industry and occupation 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 those industries that we believe account for the vast majority of informal supplier activities. As indicated in Table 1, our list contains 12 broad industry categories.

Each of the broad industry categories in Table 1 is associated with a detailed listing of specific industries and occupations. To implement our methodology, it is first necessary to develop “crosswalks” that link the detailed industry and occupation codes associated with our 12 broad categories in our tax return data source (the NRP) to the comparable industry and occupation codes used in our primary survey data source, the Current Population Survey (CPS).

For each Schedule C (self-employment) return, the NRP identifies the industry category for the business using the North America Industry Classification System (NAICS). We carefully reviewed the NAICS codes and identified those that were relevant to each of the 12 industry categories described in Table 1. Our primary source of survey information on earnings is the CPS. Unfortunately, the 2002 CPS (which corresponds to calendar year 2001, or the first NRP year) does not rely upon the NAICS.\(^7\) Rather, a different industry coding system is used. We have therefore developed a crosswalk between the relevant NAICS codes and the corresponding 2002 CPS industry codes.

Our focus is on the earned amounts of self-employment income as reported in the CPS that falls into our 12 broad informal sector industry categories, and the subsequent comparison of these amounts with NRP estimates of (unreported) informal sector income. However, past

\(^7\) As of 2003, the CPS has adopted the NAICS for industry classification.
research (Roemer, 2002) has indicated that some individuals misreport their self-employment income as wages in the CPS. It is therefore important to be able to identify reported wages that are attributable to our selected industry categories. For each primary taxpayer who reports wage and salary earnings, the NRP contains a detailed occupation code that describes the taxpayer’s main occupation. Unfortunately, no occupation code is available for the secondary taxpayer on a joint return. Later, we discuss our approach for addressing this issue. Our 2002 CPS sample also contains a detailed occupation code for wage earners; however, it is based on a different coding system. We have therefore developed a “crosswalk” between the relevant NRP and CPS occupation codes for each of our 12 industry categories, which links the relevant codes for each industry category.

To summarize, we rely on our industry code crosswalk when tabulating reported self-employment income in the CPS, and we rely on our occupation code crosswalk when adjusting for self-employment earnings that have been misreported as wages.

Note that to facilitate a comparison of our methodology for estimating the informal supplier tax gap with the approach used in earlier tax gap reports, we have also developed crosswalks between the 1986 and 2002 CPS occupation and industry codes. In addition, we have developed crosswalks between the 2002 and 2003 CPS occupation and industry codes to aid in the future development of estimates for tax year 2002 and beyond. Some examples of our crosswalks are presented in the Appendix for two of the larger industry categories: #6 (Construction) and #9 (Personal services). All crosswalks are available upon request.
3. CPS Estimates of Informal Supplier Income for 11 Industry Categories

We employ data from the Current Population Survey (CPS) to estimate earnings within 11 of our 12 broad industry categories. Note that we do not use CPS data for the food caterers 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 caterers and roadside vendors. Details of this procedure are discussed in section 4.

3.1. CPS Annual Demographic File

Our primary CPS data source is the March 2002 Annual Demographic File (ADF). 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. 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

---

8 More recent versions of this data source are now referred to as the Annual Social and Economic Supplement (ASEC).
9 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.\textsuperscript{10}

3.2. 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.\textsuperscript{11} 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

\textsuperscript{10} 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.

\textsuperscript{11} 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 were 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 were used to match individual records.\textsuperscript{12} 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 was 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.\textsuperscript{13}

3.3. Resulting CPS Measure of Reported Self-Employment Income of Informal Suppliers

\textsuperscript{12} 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 compared the recorded values of the current industry codes.

\textsuperscript{13} It was 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 were proportionally adjusted upwards to account for members of the group that were not successfully matched, either because of attrition or imperfections in the matching criteria. There was 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 proportional adjustment was applied to all matched monthly records (multiplication by 1.014) to make the weighted population total equal to the corresponding ADF total.
Here 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 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.14

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. In the case of a second job, however, it is necessary to impute earnings. Among all ADF respondents who report self-employment earnings from a second job in 2001, we find that the ratio of self-employment earnings from the second job to earnings from the longest job is 26.5 percent. To impute self-employment earnings from a second job in one of our selected industry categories (among those individuals who

---

14 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. Refer to O’Hara (2005) for a discussion of the CPS methodology for imputing potential tax filing units.
reported having a second job at the time of the 2002 monthly interview), we apply this ratio to the earnings reported for their longest job held during 2001.

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 amounted to $100.9 billion. Estimated self-employment earnings from a second job (for job changers and moonlighters combined) contributed an additional $9.2 billion, for a total of $110.1 billion.

Roemer (2002) presents evidence that some individuals in the CPS misreport their self-employment income as wages. In particular, he identifies a non-trivial 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. Averaging his tabulated results over the relevant occupations, we estimate that 4.08 percent of CPS-reported wages in our selected industry categories are actually misclassified self-employment income.

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 our 11 industry categories that are attributable to the following sources: the longest jobs held by primary and secondary filers; and second jobs held by primary and secondary filers (both job
changers and moonlighters). We do so by identifying wage-earners who are associated with the CPS occupation codes defined in our crosswalk. We then apply our estimate that 4.08 percent of the reported wages in these industry categories is actually misclassified self-employment income. The results are summarized in Table 2. Overall, we estimate that $46.3 billion in net self-employment income was 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 $156.4 billion. A comparison of this figure to the corresponding amount of net self-employment earnings actually reported on tax returns would yield our proposed measure of noncompliance.

3.4. Comparison with NRP Estimates of Net Self-employment Income

How does our methodology compare with current IRS estimates of informal supplier income, based on NRP estimates of self-employment income? In this subsection, we compare our CPS estimates of self-employment income in 11 selected informal supplier industry categories with the corresponding NRP estimates for tax year 2001. This comparison indicates that our approach generates reasonably similar estimates.

The NRP data allow us to identify members of informal supplier 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 informal supplier industries. Using this approach, we find that such taxpayers report approximately $50.9 billion in aggregate net self-employment earnings from the 11 selected informal supplier industry categories on their federal individual income tax returns. After accounting for additional self-employment income that was uncovered during the NRP examinations, overall estimated earnings amount to $86.9 billion. However, both of these measures are flawed because neither adjusts for undetected noncompliance. More relevant for comparison purposes is the NRP detection-controlled estimate of $191.3 billion, obtained by applying the estimated NRP detection-based multiplier for Schedule C income (or 3.47) for positive examiner adjustments to reported net income amounts, which corresponds roughly to the adjustment that is made by the IRS for overall Schedule C income. According to this multiplier, for every dollar in unreported net self-employment income that has been detected by an NRP examiner, another $2.47 has gone undetected. This last estimate of $191.3 billion is the NRP measure that attempts to account for undetected noncompliance, making it the “best” NRP estimate of informal supplier income.

Observe that our CPS-based estimate of $156.4 billion in net self-employment income is about 18 percent lower than the detection-controlled estimate of $191.3 billion. Our CPS results also appear to understate the overall number of informal suppliers to some extent. In particular, while the NRP data indicate that approximately 7.5 million tax year 2001 returns actually reported earnings from self-employment within the 11 informal supplier industry categories, our CPS-based estimates account for only about 7 million such returns. Still, our CPS-based

---

15 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.
16 This figure includes adjustments for cases in which self-employment income was reported on the return but on the wrong line item (e.g., when self-employment earnings was improperly reported as wages).
17 See Erard and Feinstein (2012) for a discussion of the detection-controlled estimation methodology used to produce this multiplier. Note that negative examiner adjustments were not modified by the multiplier.
estimate of net self-employment earnings vastly exceeds the $87.1 billion in net self-employment earnings that NRP examiners are able to identify through reasonably intensive audits. Thus, our methodology appears to be a useful tool for assessing the potential scope of informal supplier activities and the extent of income tax evasion associated with these activities.

After accounting for undetected noncompliance, the NRP estimates indicate that the 11 informal supplier industry categories account for about 35 percent of total Schedule C net self-employment income, but for about 42 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.

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 tax returns that report self-employment earnings from any of the 11 informal supplier industry categories is 7.5 million. Among these returns, the NRP examiners agree with this classification in 7.2 million cases (or about 95 percent of the time), meaning that the vast majority of taxpayers who self-identify themselves as operating in informal supplier industries are actually operating in these industries. On the other hand, the NRP examiners have identified an additional 1.8 million returns with self-employment earnings from one of the 11 industry categories that either have reported a code outside of these categories or have failed to report self-employment earnings.
altogether. Thus, some informal suppliers will tend to go uncounted under our methodology, making our overall estimate of activity within informal supplier industries a conservative one.

3.5. Comparison with Earlier Results

How does our methodology compare with earlier IRS estimates of informal supplier income? In this subsection, we compare results from the University of Michigan survey for 1985-1986 with results from applying an abbreviated version of our proposed methodology to 1986 CPS data for 9 comparable industry categories. Our hope is that this provides at least a rough indication of how our methodology for estimating the informal supplier tax gap compares with the one used in past IRS tax gap reports.

Excluding the food and street vendor categories of goods and services, the estimated gross receipts for informal suppliers based on the University of Michigan consumer survey amounted to $65.9 billion in 1985, including an estimated $24.9 billion in business purchases from informal vendors. The IRS assumed that net self-employment earnings were 51 percent of gross receipts in this year. Applying this assumption yields an aggregate estimated $33.6 billion in net informal supplier income.

Based on results from the same survey, McCrohan, Smith, and Adams (1991, p. 30), reported that estimated gross receipts of formal vendors from sales of the identical types of goods and services amounted to $135.1 billion in 1985. This figure does not include business purchases. If one assumes that the ratio of business purchases to household purchases was the same for formal and informal vendors, accounting for business purchases raises the estimate to

---

18 This approach implicitly assumes that the ratio of reported net income to reported gross receipts on tax returns is equivalent to the ratio of actual net income to actual gross receipts. However, this may not be the case. For example, one might expect that informal suppliers would tend to understate their gross receipts and overstate their allowable expenses on their tax returns. If so, the ratio of reported net income to reported gross receipts might be considerably smaller than the ratio of actual net income to actual gross receipts. This would imply that the net income measure used in the analysis was too low, which (in the absence of other measurement errors) would tend to make informal suppliers look more compliant than they were in actual practice.
$186.1 billion. Applying the same IRS assumption that net self-employment earnings are 51 percent of gross receipts yields an aggregate estimated $94.9 billion in net formal supplier income for 1985.

Combining these two results, aggregate estimated net income among formal and informal suppliers within the selected goods and services categories based on the University of Michigan survey approach amounted to $128.5 billion in 1985.

We have applied an abbreviated version of our proposed CPS-based methodology to estimate net self-employment earnings within 9 of our selected industry categories for 1985, categories that correspond to the goods and services covered in the above analysis. These crosswalks are described in the Appendix.

Based on the 1986 CPS, approximately $39 billion in net self-employment earnings were reported by individuals who had a longest job in one of the 9 selected industry categories in 1985. To account for self-employment earnings from a second job in one of these categories, we assume that the ratio of total earnings to longest job earnings presented in Table 2 for 2001 also applies to 1985. To account for self-employment income that has been misclassified as wages in the CPS, we assume that the ratio of our adjusted estimate to our raw estimate presented in Table 2 for 2001 also applies to 1985. After allowing for these adjustments, our estimate of aggregate net self-employment income within the 9 selected industry categories in 1985 amounts to $61.0 billion. This estimate includes earnings from both formal and informal sole proprietors in the 9 selected industry categories. While our estimate should therefore exceed the net earnings of informal suppliers in these categories, it should fall short of the combined net income of all formal and informal suppliers. This is because many of the formal sales in these industry

---

19 Our CPS analysis excludes the transportation and moving and arts and entertainment categories, as well as the caterers and roadside vendors category that is addressed in our methodology using results from the CES.
categories are presumably attributable to sales by partnerships and corporations, which are not accounted for in our estimate.

In fact, our estimate of $61.0 billion for 1985 substantially exceeds the estimated $33.6 billion in net informal supplier income based on the University of Michigan survey approach, while falling well short of the estimated $128.5 billion in combined formal and informal supplier net earnings based on that approach. It therefore appears that our estimate of net earnings falls well within a plausible range.

4. CES Estimates of Informal Supplier Income for Food Caterers and Roadside Stands

4.1. CES Estimates

As noted earlier, we are not able to use the CPS to estimate self-employment earnings for one industry category, food caterers 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 from the Consumer Expenditure Survey (CES). 20

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

---

20 Note that we use here the “Schedule C Principal Business or Professional Activity (NAICS) Codes” of:
  722300 Special food services (including food service contractors & caterers)
  445230 Fruit and vegetable markets
Also, we use the “NRP Occupation Code” of 521: Food & beverage preparation/service occupation.
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 caterers 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 caterers and roadside stands is based on the Detailed Expenditure Files (DEF) from the 2001 CES Interview Survey. Included under “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.\(^{21}\) On an annual basis, 2001 total expenditures were $4.13 billion for food caterers and $1.68 billion for roadside stands, for a combined total of $5.81 billion. See Table 3.

4.2. Comparisons with Other Estimates

It is worth noting that the above estimates of expenditures on caterers 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. Similarly, 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

\(^{21}\) 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”.

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 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 food caterers and roadside stands income 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, while our measure probably understates overall noncompliance by filers and nonfilers, it performs fairly well.

5. Summary and Conclusions

Our proposed methodology has a number of advantages over the methodology used in previous and current estimates of the informal supplier tax gap. Focusing mainly on the CPS methods, these advantages include:

- 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;
- Sales to both consumers and businesses are fully accounted for;
• 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 publicly available information is an especially compelling advantage.

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 noncompliance within the informal sector: if true earnings in this sector are understated on the survey, the difference between our survey-based estimate of “true” income and our tax return-based estimate of reported income will also be understated.

However, the evidence on this issue is encouraging. In particular, Roemer (2002) examines an exact match between earnings reported by respondents in the CPS and their Detailed Earnings Records (DER) from the Social Security Administration. He finds evidence that many respondents, particularly those employed in informal occupations, report earnings on the CPS that are not recorded in the DER, and he interprets this as evidence that the CPS measure of income includes underground earnings. Indeed, he concludes that the CPS picks up more of these earnings than another commonly used survey, the Survey of Income and Program Participation.

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 for developing countries in which informal suppliers are likely to account for a large share of economic activity – and a large amount of tax evasion. 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).

References


Appendix
Examples 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
- 233300 Nonresidential building construction
- 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

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

CPS Industry Classification

2003 CPS:
- 0770 Construction (Includes the cleaning of buildings and dwellings is incidental during construction and immediately after construction)

2002 CPS:
- 60 Construction

1986 CPS:
- 60 Construction

CPS Occupation Classification

2003 CPS:
- 0220 Construction managers
- 6200 First-line supervisors/managers of construction trades and extraction workers
- 6210 Boilermakers
- 6220 Brickmasons, blockmasons, and stonemasons
- 6230 Carpenters
6240 Carpet, floor, and tile installers and finishers
6250 Cement masons, concrete finishers, and terrazzo workers
6260 Construction laborers
6320 Operating engineers and other construction equipment operators
6330 Drywall installers, ceiling tile installers, and tapers
6350 Electricians
6360 Glaziers
6400 Insulation workers
6420 Painters, construction and maintenance
6440 Pipelayers, plumbers, pipefitters, and steamfitters
6460 Plasterers and stucco masons
6500 Reinforcing iron and rebar workers
6510 Roofers
6520 Sheet metal workers
6530 Structural iron and steel workers
6600 Helpers, construction trades
6660 Construction and building inspectors
6700 Elevator installers and repairers
6710 Fence erectors
6720 Hazardous materials removal workers
6740 Rail-track laying and maintenance equipment operators
6750 Septic tank servicers and sewer pipe cleaners
6760 Miscellaneous construction and related workers
9510 Crane and tower operators
9520 Dredge, excavating, and loading machine operators

2002 CPS:
  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

1986 CPS:
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
865 Helpers, construction trades
869 Construction laborers

Category 9: Personal Services

Schedule C Principal Business or Professional Activity (NAICS) Code
541920 Photographic services
812111 Barber shops
812112 Beauty salons
812113 Nail salons
812320 Drycleaning & laundry services (except coin-operated) (including laundry & drycleaning drop off & pickup sites)
812910 Pet care (except veterinary) services
812990 All other personal services

NRP Occupation Code
326 Photographer
503 Launderer or ironer
504 Private household cook
505 Housekeeper or butler
507 Private household cleaner or servant
509 Private household occupation
525 Personal service (barber/guide/usher/bellhop/etc.)

**CPS Industry Classification**

*2003 CPS:*
8970 Barber shops
8980 Beauty salons
8990 Nail salons and other personal care services
9070 Drycleaning and laundry services
9090 Other personal services
9290 Private households

*2002 CPS:*
761 Private households
780 Barber shops
771 Laundry, cleaning, and garment services
772 Beauty shops
791 Miscellaneous personal services

*1986 CPS:*
189 Photographers
761 Private households
780 Barber shops
771 Laundry, cleaning, and garment services
772 Beauty shops
791 Miscellaneous personal services

**CPS Occupation Classification**

*2003 CPS:*
2910 Photographers
4230 Maids and housekeeping cleaners
4320 First-line supervisors/managers of personal service workers
4350 Nonfarm animal caretakers
4500 Barbers
4510 Hairdressers, hairstylists, and cosmetologists
4520 Miscellaneous personal appearance workers
4530 Baggage porters, bellhops, and concierges
4540 Tour and travel guides
4550 Transportation attendants
4620 Recreation and fitness workers
4650 Personal care and service workers, all other
8300 laundry and dry-cleaning workers
8310 Pressers, textile, garment, and related materials
3630 Massage therapists

*2002 CPS:*
189 Photographers
403 Launderers and ironers
404 Cooks, private household
405 Housekeepers and butlers
407 Private household cleaners and servants
449 Maids and housemen
456 Supervisors, personal service occupations
457 Barbers
458 Hairdressers and cosmetologists
461 Guides
463 Public transportation attendants
464 Baggage porters and bellhops
469 Personal service occupations, n.e.c.
487 Animal caretakers, except farm
747 Pressing machine operators
748 Laundering and dry cleaning machine operators

1986 CPS:
403 Launderers and ironers
404 Cooks, private household
405 Housekeepers and butlers
407 Private household cleaners and servants
449 Maids and housemen
456 Supervisors, personal service occupations
457 Barbers
458 Hairdressers and cosmetologists
463 Guides
465 Public transportation attendants
466 Baggage porters and bellhops (best to skip – tip earner)
469 Personal service occupations, n.e.c.
487 Animal caretakers, except farm
747 Pressing machine operators
748 Laundering and dry cleaning machine operators
Table 1: Key Industry/Occupation Categories for Informal Suppliers

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

Table 2: Self-employment Income for 11 Industry Categories

<table>
<thead>
<tr>
<th>Source of Income</th>
<th>Amount ($ billions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Net Self-employment Income</td>
<td></td>
</tr>
<tr>
<td>Longest job</td>
<td>100.9</td>
</tr>
<tr>
<td>Second job:</td>
<td></td>
</tr>
<tr>
<td>Job changers</td>
<td>2.9</td>
</tr>
<tr>
<td>Moonlighters</td>
<td>6.3</td>
</tr>
<tr>
<td>Total Net Self-employment Income</td>
<td>110.1</td>
</tr>
<tr>
<td>Misclassified Self-employment Income</td>
<td></td>
</tr>
<tr>
<td>Longest job</td>
<td>45.1</td>
</tr>
<tr>
<td>Second job:</td>
<td></td>
</tr>
<tr>
<td>Job changers</td>
<td>0.5</td>
</tr>
<tr>
<td>Moonlighters</td>
<td>0.7</td>
</tr>
<tr>
<td>Total Misclassified Self-employment Income</td>
<td>46.3</td>
</tr>
<tr>
<td>Total</td>
<td>156.4</td>
</tr>
</tbody>
</table>

Table 3: Gross Self-employment Income for Food Catering and Roadside Stands

<table>
<thead>
<tr>
<th>Source of Income</th>
<th>Amount ($ billions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food caterers</td>
<td>4.13</td>
</tr>
<tr>
<td>Roadside stands</td>
<td>1.68</td>
</tr>
<tr>
<td>Total</td>
<td>5.81</td>
</tr>
</tbody>
</table>