eJournal of Tax Research

Volume 1, Number 2 2003

CONTENTS

UNSW
Faculty of Law

eJournal of Tax Research (2003) vol. 1, no. 2, pp. 93-109

Explaining the U.S. Income Tax Compliance Continuum[+]

Brian Erard[*] and Chih-Chin Ho[†]

Abstract

Within an economy, tax compliance behavior falls along a continuum. At one extreme are households who fully report and pay their tax obligations despite any opportunities or incentives to cheat. At the other extreme are households who undertake considerable efforts to conceal their income and repudiate their tax responsibilities. Using a micro-simulation database, we undertake a preliminary statistica

ISSN 1448-2398

II. DATA SOURCES

The core elements of our micro-simulation data base are derived from two separate TCMP studies that were conducted for tax year 1988, one for filers and another for nonfilers. Although these data are now some 15 years old, they have the advantage of providing detailed compliance information about both filers and nonfilers for a common tax year. We recognize that the magnitude and composition of tax noncompliance are likely to have changed since these data were collected. Notwithstanding, we believe that the data remain informative about the fundamental nature of the compliance decision and the broad underlying factors associated with noncompliance.

TCMP Filer Data

The data for filers of 1988 federal income tax returns are taken from the IRS TCMP Phase III Survey. This survey contains the results of intensive line-by-line audits of a stratified random sample of approximately 54,000 individual income tax returns for tax year 1988. For most line items both the amount that was reported by the filer and the amount that the examiner determined should have been reported are available. For income items, changes assessed by the examiner to the amount originally reported by the taxpayer are broken down according to whether the change was based on a review of third party information return documents or if it was based on other information. As discussed below in section 3, this distinction is useful for purposes of imputing additional non-detected income to taxpayer returns. A code is also available for the primary filer's occupational category based on the IRS examiner's assessment of the filer's main line of work. A set of sample weights is included to make the data representative of the national return population.[3]

TCMP Nonfiler Data

Our data on nonfilers comes from the examination-based segment of the IRS TCMP Phase IX Nonfiler Survey. The special TCMP study began with a stratified random sample of 23,283 potential nonfilers from a population of 83 million individuals for whom there was no record of a 1988 individual income tax return being filed.[4] Revenue officers set out to locate each of the individuals in this sample to determine whether they should have filed an individual income tax return for tax year 1988.[5] A total of 18,689 of the 23,283 potential nonfilers were successfully located through the search process. The revenue officers had access to information documents and past filing records. Using these records along with the information they collected during an interview or field visit with the individual, the officers made a determination whether the individual was required to file a return; i.e., whether the potential nonfiler was a "true nonfiler". Tax returns were secured from 3,546 individuals who were deemed to have been in violation of their tax filing requirements, and a random sample of 2,195 of these returns were subjected to intensive line-by-line audits, comparable to the audits performed for the TCMP Phase III study of individual return filers. It is the details from these 2,195 examined returns that we include in our micro-simulation data base. As with the filer data, the nonfiler records include the occupation of household head as well as detailed line item information about the sources and levels of household income, deductions, credits, and expenses.

Since not all potential nonfilers in the original sample of 23,283 were located, it is highly likely that a number of true nonfilers went unidentified.[6] We have therefore modified the sample weights for our sample of 2,195 located true nonfilers to make these individuals broadly representative of all true nonfilers using an econometric approach that accounts for the likelihood that a household could be located on the basis of the information (prior tax returns, third-party information slips, etc.) that was available to the revenue officer at the time he began searching for the household. This approach, which we previously employed in developing the official IRS estimate of the nonfiling tax gap, is described in Internal Revenue Service (1996).[7]

Combined Sample

To develop our core data base, we merged together the detailed information (both per return and per exam) from the TCMP filer and nonfiler data files. When weighted, our combined sample of approximately 56,000 households represents an estimated population of 112.3 million, including 104.3 million filers and 9 million nonfilers. For each household, the data base allows us to identify the occupation of the primary taxpayer and assess the sources and magnitudes of noncompliance. It also includes an imputed variable meant to approximate the burden associated preparing and filing a tax return for each of the households in our sample. This variable was defined using an IRS formula for the average time burden, in hours, for an individual whose return contains a particular set of forms and schedules.[8]

III. Imputation of Undetected Noncompliance

Even intensive examinations such as those conducted under the TCMP cannot fully uncover all noncompliance that is present. Unless undetected noncompliance is accounted for, TCMP results can provide a misleading account of the degree to which different households and occupational groups comply with their tax obligations. Below, we briefly summarize the methodology we employ to impute undetected noncompliance to returns in our micro-simulation data base. Further details are provided in Erard and Ho (2003).

TCMP examinations are generally believed to be very effective in identifying improper reports of deductions, credits, and expenses. As well, examiners have relatively little difficulty uncovering noncompliance on key income items (such as wages and interest) that are reported by third parties. For all such items, we assume that any noncompliance is fully uncovered during the examination. Our imputation of undetected noncompliance is therefore restricted to the subset of income items that not subject to information reporting. To account for undetected noncompliance, we follow a procedure similar to that employed by the IRS to generate its official estimates of the individual income tax gap—the difference between the amount of income that households owe and the amount they voluntarily pay in a timely manner.

General Imputation Approach

In most cases, we follow the IRS in assuming that for every dollar of undeclared income detected without the aid of third-party information returns, there is another $2.28 that has gone undetected by the examiner. This assumption is based on a special TCMP study conducted for tax year 1979, from which the IRS determined that examiners, on average, were able to identify only a little less than one third of undeclared income amounts when they did not have access to information returns.

Imputation of Tip Income

One major exception to this general approach for imputing undetected noncompliance is our treatment of undeclared tip income. Rather than expanding the undeclared tip income that the TCMP examiner uncovered to account for non-detection, we have replaced the TCMP examiner figure with an independent estimate of tip underreporting by the Bureau of Economic Analysis (BEA). For tax year 1988, the BEA estimated that filers reported only $5.9 billion in tips on their returns, understating their true tip income by $11.6 billion. In the absence of specific information on who understated this income, we identified some 4.4 million filers in our database that were likely to receive tip income on the basis of their occupation codes (waiters, barbers, hairdressers, bellhops, etc.), and we assigned each an equal share of the $11.6 billion (approximately $2,650 each). We employed a comparable approach to allocate $532 million in tip income to nonfilers based on the BEA estimate for nonfilers.[9]

Imputation of Informal Supplier Income

A second major exception to our general imputation approach is our treatment of "informal suppliers." The 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.

(Internal Revenue Service, 1996, p. 43)

Examples include self-employed domestic workers, street-side vendors, and moonlighting tradesmen. Conceptually, the informal economy includes all types of market economic activity that are potentially under-measured in the National Accounts owing to the vendors' informal business style (sales in cash, lack of adequate records of sales and purchases, etc.) Since the detection of noncompliance among such individuals is likely to be especially difficult, the IRS commissioned the Survey Research Center of University of Michigan to conduct some special studies during the 1980s to derive estimates the gross sales revenue earned by informal suppliers. Rather than attempt to interview the *supplier*s of goods and services in the informal economy (who might not be forthcoming about their activities), the University of Michigan researchers elected to interview the *purchasers*. Specifically, they relied on telephone wco eEar0.0gn(telepgsa7.3(me)-5.5(special studi7fuc0.21980.001(y)]TJ-15.224 -1slied tct)0.0gn6rTw population. We assumed that all of the Schedule C (self-employment) net income reported by these households ($9.5 billion by filers and $9.7 billion by nonfilers) on their tax returns was attributable to informal activities. The aggregate difference between our measures of true and reported informal supplier income for each group represented our estimate of total undeclared income. In the absence of specific information about the relative levels of unde

Overall Noncompliance by Occupation

Table 1 presents our estimates of overall noncompliance by occupational category, which accounts for both nonfiling and misreporting. On net, underpayments of tax liability more than offset overpayments within each category, so that the average level of noncompliance is positive in all cases. The occupations in the table are ranked in order of the average dollar level of noncompliance. By this measure, the 5 least compliant occupations are: (1) vehicle sales; (2) investors; (3) informal suppliers; (4) lawyers and judges; and (5) doctors and dentists.

At the other end of the continuum, the 5 most compliant occupations are: (1) the "other" occupation category, which includes homemakers; (2) military; (3) administrative support; (4) retired or disabled; and (5) production/manufacturing.

As stressed in Erard and Ho (2003), however, the compliance rankings differ when noncompliance is measured in terms of the aggregate percentage of taxes unpaid rather than the average level of noncompliance. For instance, as noted above, lawyers and judges rank fourth highest in terms of the average level of noncompliance, underpaying taxes by an estimated average of $2,273 per return. However, this represents only about 8.9 percent of their estimated overall tax liability, compared to an estimated 14.9 percent underpayment for all occupations as a whole. Similarly, doctors and dentists rank high in terms of the average estimated dollar level of noncompliance ($2,181), but low in terms of the estimated share of their overall liability that goes unpaid (7 percent).

Conversely, certain occupational groups rank relatively low in terms of average dollars of noncompliance, but quite high in terms of the aggregate share of tax liability that goes unpaid. For instance, individuals employed in service occupations other than those associated with tip earners, informal suppliers, or protective services ("other services") understate their taxes by an estimated $371 – well below the mean of $655 for the population as a whole. However, this represents some 33.1 percent of their estimated overall tax liability, which is very large relative to the average underpayment rate of 14.9 percent. Similarly, helpers and handlers (who do routine work under close supervision, such as assisting skilled workers in the construction trades, stocking grocery shelves, or packing or moving freight, cargo, or materials) are estimated to understate taxes by the relatively low amount of $409 on average, but this represents 23.8 percent of their estimated overall tax liability.

Although the relative compliance rankings for the above occupational groups depend critically on whether noncompliance is measured in absolute or percentage terms, many groups rank consistently high or low under both types of measure. For instance, the vehicle sales group ranks highest both in terms of estimated average level of noncompliance ($6,406) and estimated share of overall taxes not paid (51.1 percent). Other occupational groups that rank consistently high in terms of noncompliance are:

declaring income. Similarly, a household from an IRS district with a relatively high

Results for Filer Sample

Most econometric studies of tax compliance have relied solely on data for filers of tax returns. To investigate whether the exclusion of nonfilers from the sample leads to biased inferences, we have repeated our analysis of the variation in compliance by occupation using only the data from our filer sample. As summarized in Table 4, the filer sample results are qualitatively very similar to the full sample results in Table 3, although the regressor for the percentage of married taxpayers loses its statistical significance in the restricted sample. Thus, restricting attention to filers does not seem to impart much bias on inferences concerning the determinants of noncompliance.

VI. CONCLUSION

In this paper, we have performed a preliminary analysis of noncompliance by occupation using a micro-simulation base that contains information on both filers and nonfilers of U.S. federal individual income tax returns. We began by deriving a map of where 34 distinct occupational groups fall along the compliance continuum. The results show that, for many occupational groups, the relative ranking depends on whether compliance is defined in absolute terms or as a share of taxes owed.

Using a grouped data regression analysis, we have explored what factors are responsible for the variation in compliance along the continuum. The results indicate that opportunity plays a key role in determining which occupations are relatively compliant and which are relatively noncompliant. More specifically, compliance tends to be substantially lower among those occupations with relatively little income subject to third party information reporting. Further, noncompliance tends to increase with the time burden associated with preparing and filing a return. This may be an indication that a large burden discourages some households from filing and drives others to report dishonestly. The time burden regressor also serves as a proxy for legal ambiguity. Therefore, the result may also be an indication that ambiguity provides savvy taxpayers and tax practitioners with an improved opportunity for noncompliance, while increasing the likelihood that less able taxpayers will make unintentional errors.

Although opportunity and burden were found to play the greatest roles in explaining the variation in compliance by occupation, the percentages of elderly individuals and married couples were also significant explanatory variables. In particular, occupations with larger shares of such individuals, other factors equal, tend to be relatively more compliant.

A comparison of the grouped data regression results based on the full sample of filers and nonfilers with those based on filers alone indicates that the exclusion of nonfilers in past empirical studies may not have imparted much bias on qualitative inferences about the determinants of noncompliance. However, it is clear from the breakdown of compliance by occupation in Table 2 that one cannot fully understand compliance among such occupational groups as informal suppliers, helpers and handlers, and other services without examining both filing and nonfiling behavior.

APPENDIX - TABLES

TABLE 1: DISTRIBUTION OF NONCOMPLIANCE BY OCCUPATION, RANKED BY ESTIMATED AVERAGE LEVEL OF NONCOMPLIANCE

Occupation	Avg. level of noncompliance	% of total taxes not paid	Group's share of population	Group's share of total tax gap
Vehicle sales	$6,406	51.1%	0.1%	0.49%
Investors	$4,398	15.0%	0.2%	1.38%
Informal suppliers	$4,011	44.1%	3.0%	18.66%
Lawyers and judges	$2,273	8.9%	0.5%	1.73%
Doctors and dentists	$2,181	7.0%	0.5%	1.78%

Occupation	Avg. level of noncompliance	% of total taxes not paid	Group's share of population	Group's share of total tax gap
Post-secondary teachers	$433	6.3%	0.3%	0.19%
Other teachers, counselors, librarians	$416	10.1%	2.1%	1.31%
Helpers and handlers	$409	23.8%	7.1%	4.42%
Accountants, auditors, tax preparers	$386	5.4%	1.1%	0.65%
Other health workers	$372	10.2%	3.1%	1.74%
Other services	$371	33.1%	4.8%	2.72%
Technologists & technicians (other than health)	$344	6.7%	2.1%	1.10%
Protective services	$300	7.7%	1.6%	0.73%
Production/manufacturing	$296	9.8%	11.8%	5.34%
Retired or disabled	$281	8.8%	7.0%	3.00%
Administrative support	$176	8.0%	7.8%	2.11%
Military	$131	7.4%	1.4%	0.27%
Other	$47	8.2%	9.4%	0.67%
All occupations combined				

TABLE 2: DISTRIBUTION OF NONCOMPLIANCE BY

Occupation	**Filers**		**Nonfilers**		**Filers & nonfilers combined**	
	Avg. level of non-compliance	**% of filer popn.**	**Avg. level of non-compliance**	**% of nonfiler popn.**	**Avg. level of non-compliance**	**% of overall popn.**
Managers, consultants, public relations	$645	2.3%	$1,051	1.40%	$666	2.2%
Transportation & material moving	$538	3.0%	$2,752	0.63%	$577	2.8%
[redacted]	$554	2.8%	$1,593	0.55%	$571	2.6%

TABLE 3: RESULTS OF GROUPED DATA REGRESSION TO EXPLAIN VARIATION IN TOTAL NONCOMPLIANCE BY OCCUPATION;

REFERENCES

Alm, James (1991). “On the experimental analysis of taxpayer reporting,” *The Accounting Review,* 66:3, 577–593.

Alm, James, Roy Bahl, and Matthew N. Murray (1991). “Tax base erosion in developing countries,” *Economic Development and Cultural Change,* 39:4, 849–872.

Andreoni, James, Brian Erard, and Jonathan S. Feinstein (1998). “Tax compliance,” *Journal of Economic Literature*, 36:2, 818–860.

Erard, Brian and Chih-Chin Ho (2001). “Searching for ghosts: who are the nonfilers and how much tax do they owe?,” *Journal of Public Economics,* 81:1, 25–50.

Erard, Brian and Chih-Chin Ho (2003) “Mapping the compliance continuum,” Paper prepared for conference on “The hard to tax: international perspective,” Andrew Young School of Policy Studies, International Tax Program, Georgia State University, Stone Mountain Park, May 15-16.

Erard, Brian (1997). “Self-selection with measurement errors: A microeconometric analysis of the decision to seek tax assistance and its implication for tax compliance,” *Journal of Econometrics,* 81: 2, 319–356.

Feinstein, Jonathan S. (1991). “An econometric analysis of income tax evasion and its detection,” *RAND Journal of Economics*, 22:1, 14–35.

Internal Revenue Service (1996). “Federal tax compliance research: Individual income tax gap estimates for 1985, 1988, and 1992,” Publication 1415 (Revised 4-96), Washington, DC.

Schneider, Friedrich and Dominik H. Enste (2000). “Shadow economies: size, causes, and consequences,” *Journal of Economic Literature*, 38:1, 77–114.

Slemrod, Joel and Shlomo Yitzhaki (2002). "Tax avoidance, evasion, and administration," in *Handbook of Public Economics, Volume 3*, eds. A. Auerbach and M. Feldstein, 1423–1470.

Smith, J. and T. Adams (1987). “The measurement of selected income flows in informal markets 1981 and 1985-1986,” Survey Research Center, University of Michigan, April.

ENDNOTES

---

[1] Refer to Andreoni, Erard, and Feinstein (1998) and Slemrod and Yitzhaki (2002) for reviews of this literature. Erard and Ho (2001) provide one of the only empirical analyses of nonfilers.

[2] Unfortunately, this data base is not in the public domain, because it contains sensitive individual taxpayer information that cannot be publicly disclosed.

[3] The TCMP filer population excludes returns that were filed late as well as returns filed by non-resident taxpayers.

[4] Non-residents and individuals without valid social security numbers were excluded from the analysis.

[5] In the U.S., households with income below a specified filing threshold that varies according to age. marital, and dependency status are not required to file a federal income tax return.

[6] Unlocated individuals in the sample tended to have much larger sample weights as a consequence of the way the sample was stratified. The sample weights for the 4,594 individuals in the sample aggregate to approximately 43 percent of the potential nonfiler population.

[7] Our approach includes an enhancement to the original IRS approach in that we adjust the weights separately by sampling stratum to make the 2,195 returns broadly representative of all nonfilers who were located during the search process. For the 1996 tax gap report, the IRS adjusted the sample weights for all 2,195 returns by the same factor.

[8] We employ the IRS measure of filing burden originally developed by Arthur D. Little, Inc., which is computed by aggregating the estimated average completion times associated with each form and schedule used by the taxpayer. Thus, in essence, the measure reflects a weighted number of forms and schedules, where the weights are the estimated completion times.

[9] This estimate represents “true nonfilers”; individuals with no legal filing requirement were separately estimated to have received $93 million in tips.

[10] Our calculator ignores issues such as the Alternative Minimum Tax, but does take into account the phase-out of personal exemptions that applies to taxpayers with high levels of income.

[11] The principal difficulty was computing the additional self-employment tax for married joint filers. For such households, it was not possible using our data to determine what shares of additional self-employment and wage and salary income were attributable to each spouse. Nor was it possible to determine which households were entitled to use the optional method for computing self-employment taxes.