The Case of the Missing Strangers: What we Know and Don't Know About Non-Filers

James Cilke[*]

Abstract

Approximately 90 percent of the U.S. population "appears" annually on a Federal individual income tax return, either as the primary filer, a spouse, or a claimed dependent. Microdata reported on such tax returns, when linked to other microdata sources such as W2s, provides a wealth of economic and demographic information about the U.S. population. However, as a data source, information from tax returns has an obvious hole; it is missing information about non-filers. Simply put, a non-filer is a U.S. resident that does not appear on a Federal income tax return filed in a given year. Just because non-filers do not appear on a tax return does not mean they are "untouched" by the Federal tax system and its tax-related data sources. Most non-filers have some information about them reported to the IRS on the multitude of different Information Returns. The modeling objective is to augment a sample of tax filers with non-filers such that the microdata represents the entire U.S. population. This paper describes a measure of the non-filing population using Information Returns and discusses what we can and cannot say about non-filers.

Many government agencies and researchers use data sampled from individual tax returns to describe the income and certain demographic characteristics of the U.S. population. Arguably, the best annual source of micro-level information on the levels and sources of income in the U.S. is the sample of tax returns drawn annually by the Statistics of Income division (SOI) of the Internal Revenue Service. These data form the backbone of the individual tax simulation models used by the Joint Committee on Taxation (JCT), Treasury department, Congressional Budget Office, and other researchers both within and outside of government.

An obvious drawback to the SOI data is that the sample only covers the population who file an income tax return. The 2011 SOI sample of tax returns represents approximately 145.4 million tax returns.[1] After excluding the approximately 9.4 million returns filed by dependents and persons living overseas, the filing population accounts for an apparent 288.1 million U.S. persons. This is far short of the 315.3

million persons estimated to be residing in the U.S. during that year by the Census Bureau[2], a difference of 27.2 million people. However, as I will show later on, the size of the non-filing population is much larger than this.

Literature Overview

Over the years, a number of people have tried to quantify the non-filing population. The traditional method was to form tax filing units out of persons sampled for Current Population Survey (CPS). A subset of those CPS filing units would be statistically matched to the SOI sample of tax returns. Those residual CPS filing units not used in the statistical match formed a microdata estimate of the non-filing population.[3]

Beginning in the late 1990s, researchers began to try more direct methods of measuring the non-filing population. Cilke used a public-use file produce by the Census Bureau which exactly matched the March 1991 CPS file to a file of Federal individual income tax returns.[4] Sailer and Weber compared population estimates derived from tax administration records to those produced by the Census Bureau.[5] Mortenson, Cilke, Udell, and Zytnick constructed a file of non-filers for 2003 using Information Returns.[6] Lawrence, Udell, and Young continued on this work by building a sample of non-filers using Information Returns for 2005, and then matched those non-filers to persons in the March 2006 CPS to form tax filing units.[7] To a large degree, this paper is a continuation of the work by Mortenson et al., and by Lawrence et al.

I should also acknowledge the work of several researchers who used Information Returns and other data sources to prepare estimates of the portion of the "tax gap" generated by non-filers. For example, Erhard and Ho used data on potential non-filers from the 1998 Taxpayer Compliance Measurement Program's (TCMP) Non-filer Survey.[8] More recently, the Internal Revenue Service (IRS) used National Research Program (NRP) data in its estimates of the tax gap[9]

Data on Non-filers

Administrative data on non-filers can be derived from the multitude of Information Returns that are submitted to the IRS. For example, the Information Return type that is most familiar to people is the W-2; the statement from employers about the wages and certain benefits paid to their workers as well as amounts of tax withheld. Appendix A contains a list of the types of Information Returns received by the IRS and used in this study. The samples are constructed each year by including, for each Information Return

type, those returns issued to people with a Social Security Number (SSN) or Individual Taxpayer Identification Number (ITIN) that ends in one of ten unique 4-digit numbers. Effectively, the probability of being selected is 10 in 9999 (no person should have an SSN ending in '0000').

Taxpayers and spouses with SSNs or ITINs ending in one of the 10 unique digits are referred to as the Continuous Work History Sample (CWHS). Returns filed by people in the CWHS are automatically picked up for processing by SOI. As a result, when preparing the sample of Information Returns, SOI is generally able to determine whether the Information Return belonged to a primary taxpayer or spouse on a filed return for the same tax year.

For purposes of this study, I used only those Information Returns associated with tax year 2010. The purpose of this work was to enhance JCT's 2011-based individual tax model. As I will discuss in this paper, the additional year allows for a large number of anomalous cases to be resolved. For modeling purposes, I assume the characteristics of non-filers in 2011 look like those observed in 2010.

Next, using data available from the IRS's Compliance Data Warehouse, I obtained an individual level microdata file of the Social Security Administration's DM1 file linked to taxpayer information appearing on Federal tax returns filed between 2006 and 2011.[10] The DM1 file contains, for each person on the file, values for the year-of-birth, year-of-death, gender, a citizenship indicator, and up to 10 name controls. Further, the file contains, for each person on the file, indicators for whether that person appeared on a tax return filed between 2006 and 2011 as a primary the taxpayer, a spouse, or as a claimed dependent.[11] From this file, I obtained those records with an SSN or ITIN that ended in one of the 10 unique 4-digit numbers. I then matched these records to the Information Returns records. The result of this match is a random sample of Information Returns associated with persons who did not appear on a tax return filed for tax year 2010.

At this point, the non-filer sample consisted of 111,849 separate Information Returns associated with 41,757 unique SSNs or ITINs. Note that as the sampling rate is 10 in 9999, the sample represents approximately 41.7 million persons. The next task was to identify and eliminate records that were probably not true non-filers; i.e., "false positives" records. I did this in a series of steps, starting with most certain cases, and progressing to the least certain cases. These steps are summarized in Table 1 below.

First, some of the people in the file were simply late filers; people who appear on a tax year 2010 return that was submitted to the IRS in calendar year 2012 or later. Approximately 3.2 percent of each of SOI's annual weighted samples of tax returns are from "late" filed tax returns. SOI's "current processing year" runs from January to

December, and generally captures returns from the previous tax year. SOI's long-standing assumption is that "the characteristics of returns due, but not yet processed, can best be represented by the returns for previous income years that were processed in [the current processing year]."[12] Eliminating persons identified on late filed returns reduced the sample size by 20,173 Information Return records and 3,647 persons.[13]

Some of the observations in the file are deceased. As determined from the DM1 information, such persons died in 2009 or earlier, but are still being issued certain Information Returns. The most common Information Return types for such people are Form 1099-INT and Form 1099-DIV. Eliminating people who have died, as indicated by DM1 information, reduces the sample by 6,624 information records and 3,209 persons.

Note that the Social Security Administration is continually updating its DM1 information. There is, naturally, a lag between the time of an actual death and when that death is indicated on the DM1 file. Late filers and the lag in year-of-death information on the DM1 file are the two primary reasons why I chose to estimate the non-filing population using tax year 2010 information instead of 2011.

It is quite likely that records appear in the initial sample of non-filers because of "bad" identifier data. Fortunately, I could identify at least some of these occurrences by examining an SSN/ITIN validation code provided by IRS. Most frequently, an SSN or ITIN would be declared as "valid" if the name control from the Information Return, which consists of the first four letters of the payee's last name, matches a name control obtained from the DM1 file. IRS has a number of alternative criteria for validating an SSN. I deleted all Information Returns where the IRS has coded the SSN or ITIN as being "unmatchable." This further reduced the sample by 3,146 Information Returns and 1,538 persons.

Information Returns are not restricted to only U.S. residents; a non-resident may well receive some type of Information Return. I tried to identify and remove non-residents in two steps. First, I removed any person that received any one of five different Information Return types that would likely be issued to a non-resident. This step reduced the sample by 948 Information Returns and 391 persons. The five Information Return types are:

- Form 1042-S, Foreign Person's U.S. Source Income Subject to Withholding,
- Form TD F 90-22.1, Report of Foreign Bank and Financial Accounts,
- Form 8288-A, Withholding on Dispositions by Foreign Persons of U.S. Real Property Interests,
- Form 8805, Foreign Partner's Information Statement of Section 1446 Withholding Tax, and
- Form 4790, Report of International Transportation of Currency or Monetary Instruments (CMIR)

Second, I removed people if they receive an Information Return issued to a foreign address (not in the 50 states or District of Columbia), and the taxable income source reported on that Information Return was above a threshold of $10,000 ($1 in the case of Social Security benefits or unemployment compensation). In addition, I removed a person if all of their Information Returns had a foreign address. The goal here was to remove non-residents, including U.S. citizens living abroad, regardless of whether they have a legal obligation to file an income tax return. This step further reduced the sample by 3,901 Information Returns and 2,131 persons.[14]

In my experience, I have come to learn that one should always be a bit circumspect when working with Information Returns. In my next file-reduction step, I removed those people with large dollar amounts. I did not want to create a non-filing return created from data that is as likely to be from erroneous data as it is from an actual non-filer. Perhaps the Information Return contained mis-entered information. Perhaps the wrong SSN was entered on a filed Form 1040, and so the Information Return was never properly linked. I exclude people associated with an Information Return that contains a dollar amount in excess of a threshold. I used a $100,000 threshold to wages, gambling winnings, canceled debt, and taxable pension distributions. A $25,000 threshold applied to partnership income amounts reported on Form K1 1065, SBC income amounts reported on Form K1 1120S, and estates and trust income amounts reported on Form K1 1041. In addition, I applied a $25,000 separately, separately, to dividends, interest, rents, and royalties. Finally, I used a $1 threshold on stock received from the exercise of an incentive stock option or from an employee stock purchase plan. This step removed 1,701 Information Returns and 161 persons from the sample.

At this point, I added the population implied by the non-filer sample to the domestic population represented on tax returns, and compared these totals, by age class, to the U.S. resident population. I discovered I had a bit too many elderly people. A likely explanation for the excess elderly population is the lag in death indicators in the DM1 file. That is, the person had died in 2009 or earlier, but their death status was not reflected in the DM1 file I used at the time. I randomly dropped 3.6 percent of persons aged 71 or over from my sample. This further reduced the non-filer sample by 1,314 Information Returns and 343 persons.[15]

The measurement of the non-filing population represented by children is, by far, the most problematic. Fortunately, children represent a small fraction of the total non-filing population. Based on experience working with tax panel data, we know that SSN reporting errors on tax returns are most common for dependent children.[16] Based on a professional judgment that children in the initial sample were the most likely to be "false positives," I randomly dropped 20 percent of people aged 17 or less from the file. This

last and final step reduced the non-filer sample by 207 Information Returns and 178 persons.

Table 1 summarizes the reductions from the initial sample of apparent non-filers to the final sample used in this paper. Recall, each person represents approximately 1000 people. So, the final sample of non-filers represents approximately 30.1 million persons.

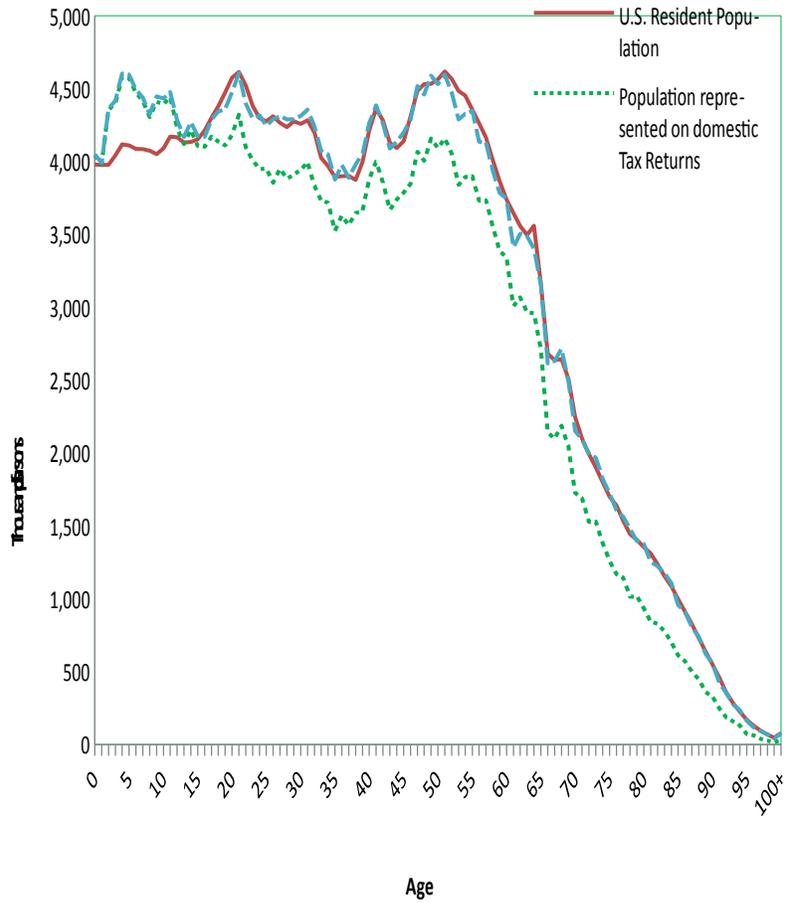| Table 1:		
Summary of the Removal of Information Returns and Persons from the Initial Sample of Non-filers to the Final Sample of Non-filing Information Returns and Persons.		
	Number	
	Information Returns	Persons
Start: Initial sample of non-filers	111,849	41,756
Less people who file late, including their dependents	-20,173	-3,647
Less people who died in 2009 or earlier	-6,624	-3,209
Less Information Returns with invalid SSNs or ITINs	-3,146	-1,538
Less people receiving "foreign-sourced" income	-948	-391
Less people with a foreign address and a large income source	-3,901	-2,131
Less people with a very large source of income	-1,701	-161
Less elderly people with imputed deaths	-1,314	-343
Less a reduction for the excess of persons under age 18	-207	-178
Equals the final sample of non-filers	73,835	30,158

At this point, I can graphically see how well the population represented by tax returns and Information Returns compares to the total U.S. resident population. In Figure 1, the solid line shows, by age, the U.S. resident population alive at any time in 2011. The dotted line shows the population on domestically filed individual tax returns. A domestic return is simply a return with a U.S. address. Finally, the dashed line shows the combined population of persons represented on filed tax returns plus the non-filer domestic population represented on one or more Information Returns.

Several observations can be made from Figure 1. For ages above 16, the combined filer and non-filer populations track the U.S. resident population amazingly well. There are a few spots where the resident population slightly exceeds the population from tax records; principally among people aged 16 to 24 and among people aged 50 to 64. We would expect that there is some portion of the population that is completely untouched by the Federal income tax system.[17]

Second, the number of children represented on filed tax returns well exceeds the U.S. resident population. Child dependents on filed tax returns with ages 0 to 15 exceed the resident population by 3.7 million persons; by approximately 4.2 million when including non-filers identified by Information Returns. The explanation for this gap is outside the scope of this paper. Some of the gap is surely due to non-compliant reporting. However, some taxpayers are legitimately claiming personal exemptions for children not living in the U.S.

Third, the number of adults between the ages of 27 and 40 for filers and non-filers is slightly greater than the U.S. resident population; by approximately 0.6 million. A likely reason which explains some of this difference is a mis-identification of overseas filers. An easy scenario to imagine is a young adult living and working overseas but using his parent's address for tax filing purposes.[18]

Figure 1:

Comparison of the 2011 U.S. Resident Population and the Population Represented on Tax Returns, by Age

Results

Table 2 shows the percent of non-filers in the data sample with each of the specific Information Return types. The rows in Table 2 are listed in order, from the most common Information Return types to the least common. The most common type received by non-filers is Form SSA-1099, which contains amounts for Social Security benefits (or the equivalent amounts of benefits paid by the Railroad Retirement Board). Here, 55.9 percent of non-filers in my sample receive an SSA-1099. Further, it is the only Information Return type received by 34.6 percent of non-filers in the sample.

Two other observations from Table 2 are worth highlighting. First 7.9 percent of non-filers (2.4 million people) receive a Form 5498, indicating they have an Individual Retirement Account. The average fair-market value of these accounts is approximately $34,000. Second, 5.0 percent of non-filers (1.5 million people) receive a Form 1098T, indicating they attend a postsecondary educational institution.

| Information Return Form Type | Information Return Form Description | Percent of all Non-filers | |
		With one or more Information Return Types	With only one Information Return Type
SSA-1099	Social Security Benefits. Includes Form RRB-1099	55.9%	34.6%
W-2	Wage and Tax Statement	24.7%	9.2%
1099-INT	Interest Income	15.6%	3.0%
1099-R	Distributions from Pensions, Retirement Plans, etc.	14.3%	0.7%
1099-G	Certain Government Payments	11.1%	3.4%
1098	Mortgage Interest Statement	9.9%	0.8%
5498	Individual Retirement Arrangement Contributions	7.9%	0.6%
1099-MISC	Miscellaneous Income	7.8%	2.8%
1098-T	Tuition Statement	5.0%	1.7%
1099-DIV	Dividends and Distributions	4.3%	0.8%
1099-B	Proceeds From Broker and Barter Exchange Transactions	2.6%	0.1%
1098-E	Student Loan Interest Statement	2.1%	0.4%
W-2G	Certain Gambling Winnings	1.2%	0.2%
1099-C	Cancellation of Debt	1.1%	0.3%
4789	Currency Transaction Report	0.7%	0.2%
K1-1065	U.S. Return of Partnership Income	0.7%	0.2%
K1-1041	U.S. Income Tax Return for Estates & Trusts	0.5%	0.1%
1099-A	Acquisition or Abandonment of Secured Property	0.5%	0.1%
1099-S	Proceeds From Real Estate Transactions	0.4%	[1]
5498-SA	HSA, Archer, Medicare & Choice MSA Information	0.3%	[1]
K1-1120-S	U.S. Income Tax Return for an S Corporation	0.3%	[1]
1099-MSA	Distributions from an MSA plans	0.2%	[1]
1099-OID	Original Issue Discount	0.1%	[1]
1099-PATR	Taxable Distributions Received From Cooperatives	0.1%	[1]
Other	All Other Information Return types.	0.3%	0.1%

Table 2:
Percent of Non-Filers with Specific Information Return Types

[1] Less than 0.05%

Table 3 shows the age and gender distributions of non-filers. The table shows that there are very few children are in the non-filing population; people aged 65 and over accounts for 38.1% of non-filers. Interestingly, non-filers under age 65 are more likely to be male.

Table 3: Counts of Non-Filer Records, by Age and Gender			
Age range	Number	Percent	Percent Male
--------------------	-----------	-----------	-----------
00 to 17	666	2.2 %	52.0 %
18 to 24	1,991	6.6 %	57.4 %
25 to 34	3,688	12.2 %	62.6 %
35 to 44	3,739	12.4 %	62.1 %
45 to 54	4,367	14.5 %	58.1 %
55 to 64	4,238	14.1 %	50.8 %
65 to 74	4,601	15.3 %	42.8 %
75 and over	6,868	22.8 %	32.0 %
Total	30,158	100.0 %	49.7 %

Table 4 shows some sundry statistics for non-filers with wages. Nearly 25 percent (7.2 million) non-filers received wages reported on a Form W-2, with wages totaling to $102.4 Billion. Of these wage earners, 5.0 million had some income tax withheld, totaling $7.4 Billion.

One frequent question is how many of these people are legally required to file. Unfortunately I cannot say precisely. An income filing threshold varies by filing status, and Information Returns do not provide information on who may be married or whether a person has dependent children. As a proxy, Table 4 splits non-filers between those with wages above and below the filing threshold for a single person. In 2011, the filing threshold for a single person was $9,500 ($10,950 if aged 65 or over) of income. Approximately 60% of the non-filers with wages (4.5 million people) had a wage value below the filing requirement threshold for a single person.

For comparison purposes, Table 4 shows the number of filers who were not required to file and had positive wages but below the filing threshold for a single person. Comparing people not required to file a return, non-filers have lower income tax withheld than filers. The average withholding-to-wages ratio is 2.6 percent for non-filers compared to 5.6 percent for filers. This is not surprising as people with a zero income tax liability have an incentive to file to claim any withheld income tax.

Table 4:
Wages and Income Tax Withholding of Non-Filers, with a Comparison to Filers

	W2 Wages		W2 Income Tax Withholding		
	Number (Mil)	Value ($Bil)	Number (Mil)	Value ($Bil)	Average Percent of Wages
All Non-Filers with Wages	7.4	102.4	5.0	7.4	4.1 %
Non-Files with Positive Wages Above the Filing Threshold for Singles	2.9	91.1	2.7	7.1	6.3 %
Non-Files with Positive Wages Below the Filing Threshold for Singles	4.5	11.3	2.4	0.3	2.6 %
Comparison group: Filers with Positive Wages below the Filing Threshold for Singles, and not required to file.[1]	10.5	53.3	8.9	3.0	5.6 %

Details may not add due to rounding.
[1] In general, a Federal income tax return is required for individuals or married couples with gross income that exceeds the standard deduction plus personal exemptions for self and spouse.

Each W-2 contains the tax identification number of each worker's employer. Using this variable, I linked each W-2 in the sample to Form 941 data. Form 941 is filed quarterly with the IRS and is used by employers to report the number of employees, total taxable wages, plus income and payroll tax withholding. More importantly for this paper, Form 941 data includes each employer's North American Industry Classification System (NAICS) code. Table 5 shows the distribution of non-filers with wages by industry. The one surprising result here is the concentration of workers in firms providing Administration and Support services. The bottom row in the Table 5 shows that 6.7 percent of non-filers and 9.1 percent of non-files with wages below the filing threshold

for singles have a 5-digit NAICS code of '56132', Temporary Help Services. Note that low-wage workers who are filers have a similar concentration in this industry class.

Table 5:
Percent Distribution of Non-filers with Wages, by Industry Class[1]

NAICS Code(s)	Industry	All	Wages Below the Filing Threshold for Singles
11	Agriculture, Forestry, Fishing and Hunting	1.5%	1.8 %
21	Mining, Quarrying, and Oil and Gas Extraction	0.4	0.2
22	Utilities	0.2	0.1
23	Construction	5.7	5.2
31-33	Manufacturing	7.1	4.8
42	Wholesale Trade	3.3	2.8
44-45	Retail Trade	9.9	11.3
48-49	Transportation and Warehousing	3.2	2.3
51	Information	1.2	0.7
52	Finance and Insurance	3.6	4.1
53	Real Estate and Rental and Leasing	1.8	1.7
54	Professional, Scientific, and Technical Services	5.1	4.5
55	Management of Companies and Enterprises	1.0	0.8
56	Administrative and Support Services, etc.	17.7	20.4
61	Educational Services	5.1	4.6
62	Health Care and Social Assistance	9.4	8.8
71	Arts, Entertainment, and Recreation	1.6	1.7
72	Accommodation and Food Services	11.8	14.0
81	Other Services (except Public Administration)	4.4	5.2
92	Public Administration	6.0	4.7
561	Addendum: Administrative and Support Services	17.4	20.2
56132	Addendum: Temporary Help Services	6.7	9.1

[1] NAICS code of the highest paying job in cases of workers with multiple employers.

Table 6 shows the distribution of observable taxable income sources and Social Security benefits for non-filers for several broadly defined income categories. Wages and Salary amounts come from Form W-2. Taxable pension and annuities, including retirement-plan distributions are reported on Form 1099-R. Unemployment compensation is reported on Form 1099-G. Capital income consists of dividends and capital gains distributions from Form 1099-DIV, interest income from Form 1099-INT, rents and royalties from Form 1099-MISC, and all taxable income sources including capital gains from the three pass-through K1 Forms (K1-1065, K1-1120S, and K1-1041). Miscellaneous income consists of canceled debt reported on Form 1099-C and gambling winnings on Form W-2G. Social Security income consists of benefits (less repayments) reported on Forms SSA-1099 and RRB-1099.

The largest observable income source for non-filers is Social Security benefits. Moreover, Social Security is the clearly dominant income source at the bottom of the income distribution; among non-filers with income under $15,000, Social Security represents nearly 80-percent of that total income. While approximately 19 percent of non-filers have some capital income, mostly dividends and interest, capital income consists of only one percent of their total income.

Information Returns likely capture the lion's share of income going to non-filers. However, I should note some obvious omissions. Topping the list are earnings from self-employment. I would note that the filing threshold for people with self-employment earnings is very low. In addition to any other filing requirements, people with self-employment earnings of $400 or more are generally required to file a Federal income tax return. And, among filers, the under-reporting of business income by individuals is the largest single component of IRS's estimate of the Tax Gap.[19]

Some sources of gross self-employment income are reported on Form 1099-MISC. However, any business expenses which would offset these gross income amounts are not reported.[20] While Forms K1-1065 and K1-1120S contain amounts for ordinary business income or losses from partnership and S-corporations, information from these forms do not distinguish whether the income or loss is "passive" or "nonpassive." For purposes of this study, I assume such income or loss is passive, and therefore is not self-employment earnings.

Similarly, capital gains are largely missing for non-filers. While Form 1099-B generally contains gross proceeds from the sale of stocks and other financial assets, the form does not contain the basis or capital gain of those assets. However, it is unlikely that capital gains are a major source of income for non-filers. As seen in Table 2 above, only 2.6 percent of non-filers receive a Form 1099-B. Note that in the future, because of recent changes to Form 1099-B by the IRS, we will be able to measure capital gains from Information Returns.

Finally, data on incomes from non-taxable sources is missing from Information Returns. However, this is true for filers as well. Such missing income sources include workers compensation, Supplemental Security Income benefits, certain Veterans' benefits, and certain other welfare benefits.

Total Income Class	Total Income		Wages and Salaries		Pensions and Annuities		Unemployment Compensation	
	Count (000)	Amount ($Bil)	Count (000)	Amount ($Bil)	Count (000)	Amount ($Bil)	Count (000)	Amount ($Bil)
$0 to $5,000	7,442	12.9	2,738	4.1	351	0.5	383	0.9
$5,000 to $10,000	6,248	47.9	905	4.8	318	0.8	438	2.6
$10,000 to $15,000	5,581	68.9	721	5.7	587	1.4	377	3.7
$15,000 to $20,000	3,599	61.9	636	7.3	918	3.0	310	4.0
$20,000 to $50,000	3,609	102.9	1,798	46.4	1,467	12.6	500	6.5
$50,000 and over	646	45.0	572	34.1	233	6.5	64	0.6
Total	27,126	$ 339.6	7,372	$ 102.4	3,873	$ 24.8	2,070	$ 18.3

Details may not add due to rounding
[1] Excludes non-filers with zero or net negative total income

	Table 6 (continued): Income from Observable Sources for Non-filers, by Total Income Class [1]					
Total Income Class	**Capital Income**		**Miscellaneous Income**		**Social Security Benefits**	
	Count (000)	Amount ($Bil)	Count (000)	Amount ($Bil)	Count (000)	Amount ($Bil)
$0 to $5,000	2,405	0.6	221	0.4	2,096	6.3
$5,000 to $10,000	574	0.6	97	0.4	5,085	38.7
$10,000 to $15,000	664	0.5	79	0.4	4,789	57.4
$15,000 to $20,000	757	0.7	63	0.3	2,933	46.7
$20,000 to $50,000	1,030	1.4	126	1.2	1,846	34.8
$50,000 and over	243	0.4	79	2.1	69	1.3
Total	5,673	$ 3.6	664	$ 4.8	16,818	$ 185.1

Details may not add due to rounding
[1] Excludes non-filers with zero or net negative total income

Summary and Conclusions

IRS's Federal income tax reporting "touches" an estimated 99.5 percent of the U.S. resident population. However, there is some uncertainty with this estimate. Approximately 90 percent of the U.S. population appears on a filed income tax return, while almost everybody else appears on one or more Information Returns filed with the IRS. In short, a microdata sample of Information Returns that are not linked with people who appear on a filed income tax return largely captures the non-filing population. In addition microdata captured from the Information Returns, when combined and linked with microdata from the Social Security Administration, provides a wealth of income and demographic information about non-filers. However, the procedure requires certain assumptions and judgment calls. And, in the end, the resulting microdata file will have its limitations. Income information for non-filers is probably very good for wages, interest, dividends, unemployment compensation, pensions and annuities, and Social Security. However, self-employment earnings are largely non-existent and certain sources of

capital gains are missing.   As for demographic information, the microdata sample does an excellent job of capturing the age and gender distribution of non-filers. However, marital status and parental status are largely non-existent.

Appendix A: Types of Information Returns		
Code	Form Name	Form Description
02	1042-S	Foreign Person's U.S. Source Income Subject to Withholding
13	TD F 90-22.1	Report of Foreign Bank and Financial Accounts
21	W-2	Wage and Tax Statement
25	3921	Exercise of an Incentive Stock Option Under Section 422(b)
26	3922	Stock Acquired Through an Employee Stock Purchase Plan Under Section 423(c)
27	5498-SA	HSA, Archer, Medicare & Choice MSA Information
28	5498	Individual Retirement Arrangement Contribution Information
30	SIIR	State Individual Information Return
31	1099-Q	Payments From Qualified Education Programs
32	W-2G	Certain Gambling Winnings
33	SCIR	State Corporate Information Return
34	SWIR	State Withholding Information Return
35	SSSTIR	State, Sales, Service or Transaction Information Return
41	8288-A	Withholding on Dispositions by Foreign Persons of U.S. Real Property Interests
46	8805	Foreign Partner's Information Statement of Section 1446 Withholding Tax
59	8596	Information Return for Federal Contracts
61	8362	Currency Transaction Report By Casinos
63	4790	Report of International Transportation of Currency or Monetary Instruments (CMIR)
64	8300	Report of Cash Payments Over $10,000 Received in a Trade or Business
65	K1-1065	U.S. Return of Partnership Income
66	K1-1041	U.S. Income Tax Return for Estates & Trusts
67	K1-1120S	U.S. Income Tax Return for an S Corporation
71	1099-H	Health coverage tax credit (HCTC) advance payments
72	5498-ESA	Coverdell ESA Contribution Information
73	1099-CAP	Changes in Corporate Control and Capitol Structure
75	1099-S	Proceeds From Real Estate Transactions
78	1098-C	Contributions of Motor Vehicles, Boats and Airplanes
79	1099-B	Proceeds From Broker and Barter Exchange Transactions
80	1099-A	Acquisition or Abandonment of Secured Property
81	1098	Mortgage Interest Statement
82	SSA-1099	Social Security Benefits. Includes RRB-1099 – Railroad Retirement Benefits
83	1098-T	Tuition Statement
84	1098-E	Student Loan Interest Statement
85	1099-C	Cancellation of Debt
86	1099-G	Certain Government Payments
89	4789	Currency Transaction Report
91	1099-DIV	Dividends and Distributions
92	1099-INT	Interest Income
93	1099-LTC	Long-Term Care and Accelerated Death Benefits
94	1099-MSA	Distributions from an Archer MSA or Medicare+Choice MSA
95	1099-MISC	Miscellaneous Income
96	1099-OID	Original Issue Discount
97	1099-PATR	Taxable Distributions Received From Cooperatives
98	1099-R	Distributions From Pensions, Annuities, Retirement Plans, IRA, Insurance Contracts

1 The SOI sample does not include tentative returns, amended returns, and certain returns that report no income. In addition, the SOI sample excludes returns obtained from IRS's Automated Substitute for Return Program.

2 The Census Bureau estimates the U.S. Resident population on July 1, for 2011 and 2012 to be 311.6 and 313.9 million persons. Using the midpoint between these two values and adding 2.5 million for residents who died in 2011 gives the estimate of 315.3 million U.S. residents who were alive sometime in 2011. See www.census.gov/popest/data/national/asrh/2012/ for Census Population estimates and www.cdc.gov/nchs/data/nvsr/nvsr61/nvsr61_06.pdf for the National Center for Health Statistics' estimate of the number of deaths.

3 For example, see James Cilke, The Treasury Individual Income Tax Simulation Model. Department of the Treasury, Office of Tax Analysis, 1994,

4 James Cilke. "A Profile of Non-Filers." U.S. Department of the Treasury, OTA working paper 78. (1998)

5 Peter Sailer and Michael Weber. "The IRS Population Count: An Update. " Washington, DC: Statistics of Income Division, Internal Revenue Service, (1998)

6 Jacob A. Mortenson, James Cilke, Michael Udell and Jonathan Zytnick. "Attaching the Left Tail: A New Profile of Income for Persons Who Do Not Appear on Federal Income Tax Returns." Washington, DC: National Tax Association, Papers and Proceedings of the 102nd Annual Conference on Taxation, 2009

7 Joshua Lawrence, Michael Udell, and Tiffany Young. "The Income Tax Position of Persons Not Filing Income Tax Returns for Tax Year 2005." In Alan Plumley ed. Recent Research on Tax Administration and Compliance. Washington, DC: IRS Research Bulletin, Internal Revenue Service. (2011)

8 Brian Erard and Chih-Chin Ho. "Searching for Ghosts: Who are the Nonfilers and How Much Tax do they Owe?" Journal of Public Economics, vol. 81, No. 1 (2001).

9 J. Russell George, "A Closer Look at the Size and Sources of the Tax Gap," Testimony before the U.S. Senate Committee on Finance, Subcommittee on Taxation and IRS Oversight, July 26, 2006.

10 This file is referred to as The IRS Databank.

11 In IRS's Masterfile data, the SSNs of up to only four dependents of are identified as appearing on a tax return.

12 Statistics of Income, 2011, Individual Income Tax Returns, Internal Revenue Service, Washington DC.

13 Nearly a year after the completion of this study, JCT received the 2012 sample of individual income tax returns. Using this updated file, I would have identified 731 additional persons in the step where I identify persons on late-filed tax-year 2010 returns. Of these, 94 would have been dropped anyway due to a subsequent sample reduction step.

14 For purposes of this study, any person on a filed tax return or an Information Return who is identified as a non-resident is assumed to be a non-resident for the entire year.

15 Several months after the completion of this study, JCT received access to an updated DM1 file. Using this updated file, I would have identified 401 additional persons in the step where I identify deceased persons. Of these, 188 would have been dropped anyway due to a subsequent sample reduction step; almost all from steps that tried to identify foreigners (which is an interesting finding in itself).

16 Some of the problems and issues of using SSNs to link tax data sets, particularly for dependent children, are discussed in Jim Nunns et. al., "Treasury's Panel Model for Tax Analysis." OTA Technical Working Paper #3, 2008.

17 For purposes of JCT's Individual Tax Model, we impute "automatons" to the population of non-filers to fully represent the U.S. resident population. Automatons are placeholders and represent the non-filing population without Information Returns. Except where noted, automatons are excluded from the analyses presented in this paper.

18 Estimates of the number of U.S. Citizens living abroad vary widely and abound with uncertainty. See Joe Costanzo and Amanda von Koppenfels, "Counting the Uncountable: Overseas Americans", Migration Policy Institute, 2013. Further, a tabulation from JCT's 2011-based Individual Tax Model shows 228,000 returns with a foreign earned income exclusion and a domestic address. In addition, 181,000 domestically filed returns received a W2 issued to an overseas address.

19 See www.irs.gov/uac/The-Tax-Gap.

20 Form 1099-MISC reports gross rental and royalty income. I assume offsetting business expenses to these gross income sources are zero.