

ATTACHING THE LEFT TAIL: A NEW PROFILE OF INCOME FOR PERSONS WHO DO NOT APPEAR ON FEDERAL INCOME TAX RETURNS*

Jacob A. Mortenson, University of Nebraska-Lincoln

James Cilke, Joint Committee on Taxation

Michael Udell, Ernst & Young

Jonathon Zytnick, Yale University

INTRODUCTION/MOTIVATION

MANY GOVERNMENT AGENCIES AND RESEARCHERS use data sampled from individual tax returns to describe the income characteristics of the U.S. population. Arguably, the best annual source of micro-level information on the levels and sources of income in the United States 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), the U.S. Department of the Treasury, 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. In 2003, roughly 130 million tax returns were filed. After excluding the approximately 10 million returns filed by dependents or persons living overseas, the filing population accounts for 259.0 million U.S. persons: far short of the 294.3 million persons estimated to be residing in the United States during that year by the Census Bureau (See Table 1). This study provides a new data set containing the income and certain demographic characteristics of individuals not represented on an individual income tax return. To that end we use a sample of information returns, including Forms W-2 and the 1099 series, not associated with individual income tax returns to construct a non-filer population and income profile.

LITERATURE REVIEW

A small body of literature has attempted to align income reporting information with the total population. An analysis by Kilss and Scheuren (1978) described the combination of Social Security

Administration (SSA), IRS, and Current Population Survey (CPS) data for 1973. These data were used to construct an earnings distribution for the United States, but only included data from those individuals with federal income tax return.

In 1998, Cilke used a public-use file of CPS micro-records matched against tax year 1991 individual tax returns. While this file included extensive information from individual income tax returns provided to the Census Bureau, some of the CPS records were not linked to an individual tax return. Cilke used these records to create a profile of the non-filing population.

Researchers at the Statistics of Income Division (SOI) of the IRS have produced two studies which include information returns on individuals who did not file federal income tax returns. The first SOI study by Sailer, Weber, and Yau (1993), used tax year 1989 federal income tax return data coupled with various information returns to construct a population estimate comparable to the Census. Sailer and Weber (1998) updated the 1993 paper with similar data from tax year 1993.

The focus of the two SOI studies, unlike the research for this paper, was to build a reasonable sample of the count of the U.S. resident population using individual income tax or information returns. In contrast, the focus of research for this paper is to create a representative sample and income profile of the non-filing population using information returns.

DATA AND METHODOLOGY

This research primarily uses two datasets from SOI. The first is a stratified random sample of tax returns representing the tax filing population for 2003. The sample of individual tax returns provided by SOI contains every filed tax return where the last four digits of the SSN for either the primary taxpayer or the spouse match one of five 4-digit sequences. Statistics of Income also obtained the information returns for any SSN where the last four

*Views in this paper are those of the authors and do not necessarily represent official policy positions of the Joint Committee on Taxation, Ernst & Young, the University of Nebraska-Lincoln, or Yale University.

Table 1
Determining the size of the 2003 Non-filing Population

	<i>Population (Mil)</i>
End-of-year 2003 U.S. Resident Population [1]	291.9
Total deaths in 2003 [2]	2.4
Total Resident Population alive at any time in 2003	294.3
Population in 2003 represented on U.S non-dependent individual tax-returns [3]	260.6
Less population on tax returns living overseas	-1.6
Total Resident Population represented on a tax return	259.0
Implied non-filing population	35.3

[1] U.S. Census Bureau (2008). The value is the Total U.S. Resident population on 1/1/2004. After the completion of this study, the Census Bureau updated its January 1, 2004 Resident Population estimate. The new estimate is 291.65 million.

[2] Centers for Disease Control and Prevention (2003).

[3] Tabulations from the 2003 Statistics of Income INSOLE file. Population consists of filers, spouses on joint returns, and any persons claimed as dependents.

digits matched one of the five sequences. A subset of the second file consists of a random sample of persons that did not appear on an individual income tax return for tax year 2003, either as a primary or secondary, but had some type of information return filed with the IRS. This file has its origins with the Continuous Work History Sample created by the Social Security Administration for their research purposes.

All information returns contain a core set of information including the tax year, the payee's SSN or Taxpayer Identification Number (TIN), and a SSN Validity Indicator. Before the JCT receives an information return, the IRS undergoes an extensive verification process concerning an individual's name and SSN. The most common validity check occurs by matching a name obtained from the information return with a name obtained from the Social Security Administration's (SSA) name and age control file referred to as the DM-1 file. A name control is simply the first four letters of a person's last name. In this study, we only used those information returns with a valid SSN as identified by the validity indicator.

The non-filer dataset for tax year 2003 is created from 17 different information returns (listed in Appendix A) with an SSN that could not be linked to a filed tax return. As there are 9,999 possible 4-digit ending numbers (no person is issued an SSN ending in 0000), the selection probability

for being included in the sample is 5 in 9,999; the weight for each non-filing record is therefore approximately 2,000. The non-filing records are then exactly linked to SSA's DM-1 file to obtain gender, date of birth, and date of death.

Prior to making any adjustments, the population associated with information returns not linked to a tax return included some 53 million people. A quick glance at Table 1 above shows that with 259 million resident persons accounted for on filed tax returns and 53 million persons accounted for who were not on a tax return, the resulting 312 million persons substantially exceeds the Census estimated population of the United States for 2003 of 294 million persons. Several adjustments are subsequently performed with the intent of getting at the true non-filing population. Individuals who died prior to 2003, dependents of filers, overseas non-filers, and those who filed their 2003 tax returns in 2004, 2005, and 2006 are removed. The resulting file consists of 29.9 million people. These 29.9 million non-filers account for 85 percent of the 35.3 million non-filers for 2003 shown in Table 1. When combined with the 259 million resident persons accounted for on filed tax returns, the federal tax system contains information on 98.2 percent of persons who were U.S. residents during 2003.

Of those, more than half (16 million) have Form 1099-SSA, suggesting a large portion of non-filers are elderly or disabled. Approximately 9.8, 8.8, and

6.2 million individuals have Forms W-2, 1099-R, and 1099-INT, respectively. Fewer than 3 million non-filers have Forms 1099-G, 1099-DIV, and 1099-MISC, while approximately one million non-filers have Forms 1098-T and 1099-B. All other forms belong to less than one million people each, with the majority of individuals having multiple forms.

Our study employs a slightly different set of data than the SOI studies largely as a result of the tax code's evolution in the past decade. The most recent SOI study used four forms not included in this study and excluded six used here.¹ For example, Forms 1098-E relating to the payment of student loan interest and 1098-T relating to attending at an institution of higher learning, did not exist at the time of the earlier SOI studies but are now valuable for gathering information on a population with many non-filers: college students.

AGE DISTRIBUTIONS OF FILERS AND NON-FILERS

The age distributions of the non-filer population with information returns and the filer population are strikingly different, especially for the young and old. Notably, 40 percent of the non-filer group is 65 or older compared to 10 percent in the filing population. This is likely driven by the weak incentives to file federal income tax returns facing older Americans. Individuals 65 years of age and older are less likely to have income withholdings as relatively few have sources of income requiring withholdings. They are also less likely to exceed the income filing threshold for a federal income tax return. Without these two incentives, which are strong for persons to whom they apply, older individuals have markedly fewer compelling reasons to file a federal income tax return.

At the other end of the age spectrum, 37 percent of the population appearing on a federal income tax return is younger than 25 compared to only 10 percent of non-filers. This is primarily because dependents of filers are included in the filer population: dependents comprise 86 percent of the filer population younger than 25. Only those dependents whose parents are also non-filers could possibly enter the pool of non-filers used here. Since dependents are less likely to have jobs, interest earning savings accounts, or other sources of income they are also less likely to have a connection to the tax world apart from their parents. Additionally, the Earned Income Tax Credit (EITC) encourages

individuals with children to file tax returns, further reducing the likelihood that dependents would not be connected to a tax return.²

The 29.9 million non-filing persons with information returns and the 259.0 million U.S. residents represented on non-dependent individual tax returns in 2003 compose the total tax system population of 288.9 million people. This represents 98.2 percent of the target population, leaving the tax population short some 5.4 million people. Figure 1 shows the distribution, by age class, of the U.S. resident population represented on either a filed tax return or an information return. The largest shortfalls are predictably found in the age group least likely to be connected to the tax system: those 24 and younger.³

A NEW PROFILE OF NON-FILER INCOME

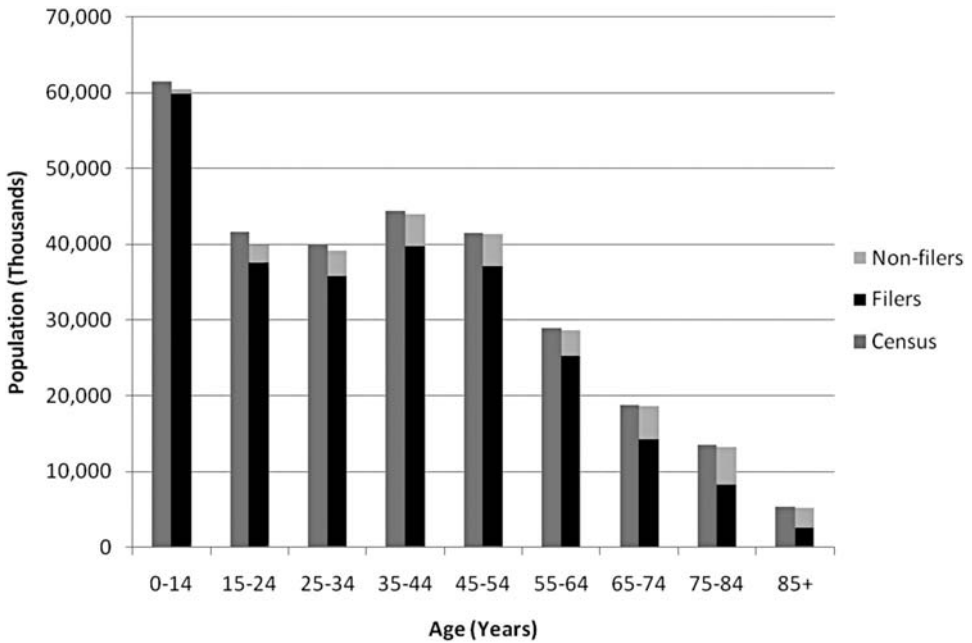
After creating a sample of persons represented in the tax system, the next step is to create a profile of non-filer income using income data found on information returns. Four categories of income are created: labor, miscellaneous, capital, and retirement income.

Labor income is primarily derived from Medicare wages on the Form W-2, but also includes unemployment compensation from Form 1099-G and nonemployee compensation, medical payment, fishing income, golden parachute, and attorney fees from the Form 1099-MISC (see Appendix A for a description of all forms used in this study). Miscellaneous income contains the amount of debt cancelled on the Form 1099-C and "other income" on Form 1099-MISC.

Capital income is found on Forms K-1, 1099-B, 1099-DIV, 1099-G, 1099-INT, 1099-MISC, and 1099-PATR. Specifically, capital income includes capital gains, ordinary dividends, non-cash liquid dividends, and cash liquid dividends on Form 1099-DIV, interest and savings bonds on Form 1099-INT, and patronage and non-patronage dividends on Form 1099-PATR. Capital income also includes agricultural subsidies and taxable grants from Form 1099-G as well as rent, royalties, substitute payments for dividends, and crop insurance from Form 1099-MISC.

Forms K-1 and 1099-B require special attention. The Form K-1 is filed by a partnership, S corporation, trust, or estate on behalf of beneficiaries and includes the following capital income items: business income, dividends, guaranteed payments,

Figure 1: Census v. Tax System Populations by Age



interest, long-term capital gains, ordinary income, other rental income, passive income, royalties, real estate, and short-term capital gains. K-1 income is the one source of income that could be negative.

Four items on Form 1099-B reporting gross proceeds also contribute to capital income: barter income, bond income, realized profit and loss, and a portion of gross proceeds. Bond income is derived from bond tax withheld, assuming an average tax rate of 25 percent. Taxable elements of gross proceeds are estimated by applying the ratio of taxable gains to gross proceeds as reported in Sales of Capital Assets (SOCA) data (only returns with less than \$250,000 of adjusted gross income are included). Using this ratio of \$.1707 per \$1 of gross proceeds, \$11.7 billion of the \$68.7 billion of gross proceeds are considered to be income.

Retirement income is composed of Social Security benefits on Form 1099-SSA and the taxable amount on Form 1099-R. Our definition of “taxable amount” on the Form 1099-R excludes rollovers to IRAs and tax-sheltered annuities, distributions from ROTH IRAs in the first five years, distribution from a ROTH IRA qualified distribution, and ROTH IRA distributions not subject to penalty.

Table 2 below presents these four aggregations of income from the information returns for each person by the total amount of income associated with that person. In total, we estimate that there was \$396.3 billion of income associated with these 29.9 million persons.

The average income of the non-filer population is a little over \$13,000. The two primary sources of non-filer income are labor and retirement income, accounting for \$194.7 billion and \$169.1 billion of the \$396.3 billion of total income. Capital and miscellaneous income are small in comparison, representing only \$28.4 billion and \$4.1 billion, respectively. In addition to having the most income in absolute terms, labor income is the largest per capita with average earnings of \$17,038. Retirement income, the most populous category, contains \$9,957 of income per person. The smaller income types have similarly smaller per capita totals, with \$2,684 for capital income and \$6,666 for miscellaneous income.

Table 2 is perhaps best summarized in terms of income categories: low, middle, and high. Non-filers with less than \$5,000 of annual income (low income) comprise roughly 32 percent of the non-filer population and account for less than 4 percent

Table 2
Total Income and Number of Non-filers by Income Type, 2003

Total Income	Labor		Capital		Retirement		Miscellaneous		Total Income	
	Count (Thou)	Amount (\$Bil)	Count (Thou)	Amount (\$Bil)	Count (Thou)	Amount (\$Bil)	Count (Thou)	Amount (\$Bil)	Count (Thou)	Amount (\$Bil)
Less than or equal \$0	12	\$0.1	56	(\$1.0)	2	\$0.1	-	-	1,154	(\$0.8)
\$0 to \$5,000	3,954	\$6.9	2,702	\$0.9	2,720	\$7.9	146	\$0.2	8,479	\$15.9
\$5,000 to \$10,000	1,700	\$10.4	2,002	\$1.1	6,075	\$43.6	106	\$0.3	7,489	\$55.4
\$10,000 to \$15,000	1,206	\$12.0	2,312	\$1.8	4,332	\$50.0	88	\$0.3	5,215	\$64.2
\$15,000 to \$20,000	958	\$14.0	1,490	\$2.3	2,112	\$32.5	66	\$0.2	2,856	\$49.1
\$20,000 to \$25,000	764	\$15.2	628	\$1.2	786	\$14.6	24	[Z]	1,398	\$31.0
\$25,000 to \$50,000	1,870	\$58.8	788	\$3.6	690	\$13.5	118	\$0.6	2,230	\$76.5
\$50,000 and over	964	\$77.2	600	\$18.4	266	\$6.9	70	\$2.4	1,078	\$104.9
Total	11,427	\$194.7	10,577	\$28.4	16,982	\$169.1	618	\$4.1	29,899	\$396.3

Details may not add due to rounding.

[Z] – Less than \$50 million

of the income. Fewer than half of these individuals have labor income and roughly one-third have retirement or capital income. Additionally, 11 percent (over 1 million) do not appear in any of the four categories. These individuals either do not have any income reported in the fields used to measure income or are included via information returns with no income fields such as Forms 1098-E or 1098-T.

The characteristics of low earners is in stark contrast to the middle income group – those with at least \$5,000 but less than \$20,000 of annual income – over 80 percent of whom have retirement income. Twenty-three percent of middle income non-filers have labor income and 37 percent have capital income. Retirement income similarly accounts for approximately 75 percent of all income for this group, while 21 percent is labor income.

Non-filers earning \$20,000 or more, the high-income group, have markedly different characteristics when compared to low- and middle-income earners. The average laborer in this group earns \$42,043, while the average capital and retirement income is \$11,506 and \$20,055, respectively. Labor income is far and away the largest source of income, accounting for 71 percent of total high-earner income (and 38 percent of all income, from all categories and income groups). Retirement and capital income are a lesser proportion, comprising 16 percent and 11 percent of total income, respectively.

Table 3 displays the same income categories of Table 2 by age. The non-filer income distribution by age is a bimodal and is skewed towards older age groups. Labor and miscellaneous incomes are skewed towards younger individuals when compared to the distribution of total income. Capital income is distributed more or less evenly for those 35 to 44 and 55 and older, but is markedly larger for those 45 to 54. Retirement income is (unsurprisingly) earned by the elderly, with 72 percent belonging to those 65 and older. Social Security benefits (disability or retirement) account for 76 percent of total income for those 65 and older. The similar figure is 14 percent for those under 65.

By way of comparison, Tables 4 and 5 show the distribution of income for people who appear on a filed tax return. Here, income amounts for the primary and secondary taxpayers are largely from amounts reported on the return's Form 1040.⁴ Income amounts for dependents appearing on the return are obtained from information returns.

Combining Tables 2 and 3 with Tables 4 and 5, respectively, would provide a measure of total income in the United States.

TAXES PAID

One of the benefits from using information returns filed by a third party on behalf of a person is that amounts withheld for payments of taxes are reported. Most individuals – specifically, those receiving a Form W-2 – pay taxes on income throughout the year via employers or other disinterested third parties depending on the type of income they earn and the number of exemptions claimed on the Form W-4. The number of exemptions an individual claims is related to several factors, including the taxpayer's prior year tax liabilities.

Table 6 displays taxable income and tax withheld for non-filers by total income. Taxable income as defined here differs from positive income in one sense: Social Security benefits are excluded. Using the JCT's individual tax model to build a profile of non-filer Social Security beneficiaries, we conclude a trivial amount of Social Security income exceeds the threshold over which benefits are subject to income taxes, and as a result, we have excluded Social Security benefits. We may revisit this assumption at a later date when additional information concerning income is collected.

In 2003, non-filers in this data set paid nearly \$15.9 billion of federal income taxes through withholdings on \$255.3 billion of taxable income. That equates to an effective tax rate of 6.2 percent. As shown in Table 6, the percentage of income withheld steadily increases with the amount of total income earned, reflecting the progressivity of the federal income tax code. Individuals with at least some tax withheld who earned less than \$15,000 withheld 5 percent of their income or less while those earning more than \$50,000 paid roughly 13 percent.

Note, however, only 8.3 million non-filers have any income withheld; some 21.6 million have no tax withheld. The income of these two subpopulations are starkly different: non-filers with some tax withheld earned \$170.8 billion of taxable income (\$20,578 per person) while the 21.6 million individuals with no tax withheld earned only \$84.5 billion (\$3,911 per person). The distribution of those without withholdings is heavily skewed toward lower income individuals: nearly 17.7 million (80 percent) earned less than the mean and only 1.5

Table 3
Total Income and Number of Non-filers by Age, 2003

Age	Labor		Capital		Retirement		Miscellaneous		Total Income	
	Count (Thou)	Amount (\$Bil)	Count (Thou)	Amount (\$Bil)	Count (Thou)	Amount (\$Bil)	Count (Thou)	Amount (\$Bil)	Count (Thou)	Amount (\$Bil)
0 to 14	42	\$0.2	208	\$0.2	322	\$1.4	14	[Z]	542	\$1.7
15 to 24	1,808	\$10.2	316	\$0.3	290	\$1.5	32	\$0.1	2,368	\$12.0
25 to 34	2,512	\$35.0	574	\$2.0	400	\$2.1	74	\$0.3	3,372	\$39.5
35 to 44	2,850	\$58.5	1,002	\$2.5	984	\$7.6	148	\$1.0	4,192	\$69.7
45 to 54	2,446	\$59.5	1,328	\$11.7	1,456	\$14.0	138	\$1.9	4,220	\$87.1
55 to 64	1,154	\$25.8	1,208	\$3.3	2,024	\$21.1	98	\$0.5	3,354	\$50.7
65 to 74	390	\$4.7	1,758	\$2.5	4,234	\$44.0	42	\$0.1	4,398	\$51.2
75 to 84	186	\$0.6	2,616	\$3.5	4,818	\$51.8	56	\$0.1	4,910	\$56.0
85 and older	40	\$0.1	1,568	\$2.5	2,456	\$25.6	16	[Z]	2,546	\$28.3
Total	11,427	\$194.7	10,577	\$28.4	16,982	\$169.1	618	\$4.1	29,899	\$396.3

Details may not add due to rounding.

[Z] – Less than \$50 million

Table 4
Total Income and Number of Filers by Income Class and Income Type, 2003

Total Income	Labor		Capital		Retirement		Miscellaneous		Total Income	
	Count (Thou)	Amount (\$Bil)	Count (Thou)	Amount (\$Bil)	Count (Thou)	Amount (\$Bil)	Count (Thou)	Amount (\$Bil)	Count (Thou)	Amount (\$Bil)
Less than \$0	681	\$7.1	2,639	(\$48.6)	367	\$2.4	807	\$0.7	2,641	(\$38.3)
\$0 to \$5,000	15,037	\$35.7	13,590	\$1.3	3,320	\$8.5	2,890	\$1.4	26,639	\$46.7
\$5,000 to \$10,000	13,267	\$93.2	7,186	\$5.5	4,618	\$29.4	1,556	\$0.9	17,231	\$129.0
\$10,000 to \$15,000	13,539	\$153.7	8,707	\$10.9	5,403	\$51.5	1,693	\$1.0	17,358	\$217.1
\$15,000 to \$20,000	13,760	\$215.2	9,866	\$16.0	6,110	\$77.0	2,023	\$1.1	17,688	\$309.3
\$20,000 to \$25,000	13,717	\$274.0	10,403	\$18.0	6,377	\$97.1	2,479	\$1.4	17,383	\$390.5
\$25,000 to \$50,000	45,521	\$1485.2	36,170	\$82.8	16,675	\$308.1	14,339	\$8.6	53,067	\$1884.8
\$50,000 and over	32,283	\$2881.0	30,204	\$648.6	9,286	\$208.6	16,545	\$27.2	34,699	\$3765.4
Total	147,805	\$5145.1	118,764	\$734.4	52,156	\$782.7	42,333	\$42.4	186,705	\$6704.6

Details may not add due to rounding.

[Z] – Less than \$50 million

Table 5
Total Income and Number of Filers by Age and Income Type, 2003

Age	Labor		Capital		Retirement		Miscellaneous		Total Income	
	Count (Thou)	Amount (\$Bil)	Count (Thou)	Amount (\$Bil)	Count (Thou)	Amount (\$Bil)	Count (Thou)	Amount (\$Bil)	Count (Thou)	Amount (\$Bil)
0 to 14	842	\$8.0	3,604	\$0.5	1,582	\$7.2	—	—	5,720	\$15.7
15 to 24	22,172	\$242.2	6,318	\$6.7	1,572	\$6.9	316	\$0.3	24,212	\$256.1
25 to 34	31,932	\$964.8	16,554	\$31.4	2,674	\$9.0	6,364	\$4.1	33,776	\$1009.2
35 to 44	35,239	\$1462.8	24,480	\$116.1	4,248	\$24.8	11,816	\$9.3	38,244	\$1613.1
45 to 54	32,409	\$1510.6	26,486	\$168.2	6,337	\$57.3	12,111	\$11.7	35,961	\$1747.9
55 to 64	18,583	\$800.2	19,842	\$170.2	11,815	\$179.3	7,577	\$10.2	24,416	\$1159.8
65 to 74	5,272	\$134.3	11,932	\$119.2	13,408	\$276.7	2,801	\$4.7	13,780	\$534.9
75 to 84	1,224	\$19.9	7,225	\$88.4	8,021	\$171.5	1,085	\$1.7	8,071	\$281.4
85 and older	131	\$2.2	2,323	\$33.7	2,499	\$50.2	262	\$0.4	2,525	\$86.5
Total	147,805	\$5145.1	118,764	\$734.4	52,156	\$782.7	42,333	\$42.4	186,705	\$6704.6

Details may not add due to rounding.

[Z] – Less than \$50 million

Table 6
Income Subject to Tax by Withholding Status and Income for Non-Filers

Total Income	With Income Tax Withholdings			Without Income Tax Withholdings			With Payroll Tax Withholdings		
	Count (Thou)	Withheld (\$Bil)	Income (\$Bil)	Withholding / Income	Count (Thou)	Income (\$Bil)	Count (Thou)	Withheld (\$Bil)	Payroll Income (\$Bil)
Less than or equal \$0	16	[Z]	[Z]	-7.94%	8,341	(\$1)	8	[Z]	\$0.1
\$0 to \$5,000	2,328	\$0.2	\$5.2	4.22%	9,989	\$12.0	3,568	\$0.4	\$5.6
\$5,000 to \$10,000	1,106	\$0.4	\$8.0	4.75%	1,746	\$12.3	1,266	\$0.6	\$7.9
\$10,000 to \$15,000	946	\$0.5	\$10.4	5.02%	558	\$6.7	876	\$0.7	\$9.2
\$15,000 to \$20,000	754	\$0.7	\$12.2	5.66%	248	\$4.3	712	\$0.8	\$11.0
\$20,000 to \$25,000	672	\$1.0	\$14.1	6.97%	134	\$3.0	622	\$1.0	\$12.7
\$25,000 to \$50,000	1,696	\$4.5	\$55.1	8.10%	300	\$10.4	1,582	\$3.7	\$49.3
\$50,000 and over	780	\$8.6	\$66.0	13.04%	286	\$36.6	760	\$3.6	\$56.4
Total	8,297	\$15.9	\$170.8	9.29%	21,602	\$84.5	9,393	\$10.7	\$152.2

Details may not add due to rounding.

[Z] – Less than \$50 million

million (7 percent) earned more than \$10,000. The withholding population is similarly skewed but to a lesser degree: 65 percent earned above the mean.

The Form W-2 accounts for the great majority of income withheld: \$14.4 billion of the \$15.9 billion of federal income tax withheld from non-filer incomes are found on Form W-2. Non-filers also paid \$10.7 billion of Federal Insurance Contributions Act (FICA) and Medicare taxes on \$152.2 billion of wage income. The percentage of wage income withheld has a slightly negative correlation with income earned, reflecting the regressivity of FICA taxes (earnings subject to taxation were capped at \$87,000 in 2003).

CURRENT POPULATION SURVEY COMPARISON

A conventional method for estimating the characteristics of the non-filing population by tax policy analysts is to derive the tax filing population from the total population represented in the Current Population Survey (CPS) produced by the Census Bureau. This is accomplished by creating tax filing units from CPS family and household records. Married couples are combined to form joint returns and children are linked to their parents. The resulting file is then divided into two groups. The first group consists of CPS tax units expected to have filed a tax return. Typically, the filing units are derived by subjecting CPS income amounts to income amounts required to file a tax return. This group is then statistically matched to an existing individual tax returns cross-sectional file.

The second group, often simply a residual, consists of CPS tax units not expected to have filed a return. The CPS sampling frame only covers the noninstitutional population of the United States. So, the last step in building a representative sample of non-filers is to augment the file with non-filers living in prisons or nursing homes. In general, no income is imputed to non-filing prisoners, while non-filing nursing home residents are imputed Social Security income and pension income. Except for these two groups, the income and demographic profile of the non-filers comes from the CPS. One benefit from this approach is that the resulting tax model file has the same number of families, including family size and age structure, as the CPS.

Table 6 shows the income distribution of non-filers at 2003 levels as used by the Joint Committee on Taxation (JCT) in their individual income

tax micro-simulation model. Note that counts in the table are nondependent individuals and – in the case of joint returns, capital, retirement, and miscellaneous incomes – are allocated equally to each spouse.

Comparing the income profile of non-filers seen in Table 7 to the income profile of non-filers derived from information returns seen in Table 2 yields several interesting observations. First, both tables have individuals with incomes well above the filing thresholds. However, in the case of the CPS non-filers, this was by design. Analysts at the JCT know that many persons file tax returns that do not have a filing requirement: either out of habit or to receive a refund. At the same time, there are persons who should file a tax return but do not. To account for this later group, a small portion of tax returns constructed from CPS information that should file a return are included in the non-filer group.

Second, much fewer CPS non-filers have labor income compared to the non-filers identified in this study. When creating filing and non-filing units out of CPS observations, the conventional approach has been to assume that people with earnings were likely to be filers, even though they may have income below the filing threshold. The presence of earnings allows people to claim the refundable earned income tax credit. Further, income taxes are often withheld from wages throughout the year; a person needs to file a return to claim a refund of the withheld taxes. However, this study of non-filers derived from information returns suggests that some people are, for whatever reason, willing to “leave money on the table.”

Earned income for the CPS non-filers includes both wages and net self-employment earnings. Unfortunately, net self-employment income is not readily available from information returns. Arguably, this is the most significant weakness of using information returns to construct the income profile of non-filers. At the same time, it is also the largest weakness of using filed tax returns since self-employment income is the most understated type of income on filed returns. Information returns do contain components of self-employment income. For example, amounts paid to an independent contractor may appear on a 1099-MISC information return. However, this is only likely to be a small portion of the gross income received by independent contractors. Further, the value on the 1099-MISC would be a part

Table 7
CPS-Derived Non-filers: Total Income from Taxable Sources on the CPS (Individuals w/Non-zero Inc)

Total Income	Labor		Capital		Retirement		Miscellaneous		Total Income	
	Count (Thou)	Amount (\$Bil)	Count (Thou)	Amount (\$Bil)	Count (Thou)	Amount (\$Bil)	Count (Thou)	Amount (\$Bil)	Count (Thou)	Amount (\$Bil)
Less than or equal \$0	0	-	0	—	0	—	0	—	0	—
\$0 to \$5,000	203	\$0.2	51	[Z]	1,256	\$4.3	0	—	1,397	\$4.5
\$5,000 to \$10,000	284	\$1.5	19	[Z]	6,207	\$47.8	0	—	6,391	\$49.3
\$10,000 to \$15,000	668	\$8.1	30	[Z]	7,116	\$85.3	0	—	7,785	\$93.4
\$15,000 to \$20,000	720	\$11.9	46	[Z]	1,304	\$21.7	0	—	2,010	\$33.6
\$20,000 to \$25,000	553	\$11.9	41	[Z]	184	\$4.0	0	—	733	\$19.5
\$25,000 to \$50,000	816	\$26.4	112	\$0.1	52	\$1.4	0	—	864	\$27.9
\$50,000 and over	331	\$29.4	117	\$0.2	20	\$1.0	0	—	351	\$30.7
Total	3,364	\$89.5	417	\$0.3	16,139	\$165.4	0	—	19,532	\$255.3

Details may not add due to rounding.

[Z] – Less than \$50 million

of the independent contractor's gross income and would not reflect any legitimate expenses incurred to produce the income. As a result, the omission of self-employment income likely indicates that even the information-returns-based income profile presented in this paper understates the amount of incomes by persons who do not file tax returns.

A third observation regards the comparison of retirement income for CPS non-filers and non-filers with information returns: the counts and amounts of retirement income are both (unsurprisingly) very similar. The largest source of income in this category is Social Security benefits. Social Security benefits are well defined and are therefore likely to be well-reported on the CPS. Further, tax withholding of Social Security benefits is quite rare and benefits are generally not subject to tax unless the person has other substantial sources of income. As a result a person with most of their income from Social Security is unlikely to file a tax return.

A fourth observation is that CPS non-filers have little capital income and no miscellaneous income. The only sources of capital income reported on the CPS are dividends, interest, and rents; conspicuously missing is income from capital gains. The CPS does have taxable income sources which would be included in the miscellaneous income (e.g., alimony). However, in constructing non-filing units from CPS records for this particular tax model, none of the designated non-filers had a miscellaneous income source. Contrast this to the non-filers from information returns: several information return types report capital or miscellaneous income sources.

DISCUSSION

This study presents a novel profile of income for persons who do not appear on filed income tax returns. Two important conclusions can be drawn from the data: the vast majority of the resident population in 2003 had information recorded with the U.S. Department of the Treasury and non-filers with information returns have much more income than other measures indicate.

This data suggests some 29.9 million of the 35.3 million persons who did not appear on a filed tax return for 2003 none-the-less had information return(s) filed on their behalf. Coupling this population with the tax return filing population reveals that 98.2 percent of the 294.3 million residents of the United States in 2003 were recorded in information

filed with the Treasury. Judging by the differences between the total tax system age distribution and the age distribution of the population prepared by the Census Bureau, the remaining 5.4 million persons not found in the tax system were likely mostly young children.

Non-filers, the 29.9 million persons with information returns, had roughly \$400 billion of income reported on information returns, \$181 billion of which was associated with 3.3 million persons who had reported amounts in excess of \$25,000. The largest area of discrepancy between this and other distributions of non-filer income, such as those generated from the CPS, is wage income: the income found on Form W-2 for non-filing persons dwarfs wage income of non-filers derived from studies that use the CPS for income distribution measurements.

Tax system information is not perfect, however. We recognize the possibility that an information return may have an incorrect SSN. As previously mentioned, we minimize this risk by only using information returns with a valid SSN Indicator. However, even in this case, a valid SSN on an information return may not represent the activity of that person if their SSN was stolen or otherwise used by another person. Identity theft (or sharing) of SSN's and names is an attractive economic crime because the person using the information will appear "valid" to many detection regimes. These types of situations are difficult to control in the information return data.

A second problem with using information returns is that certain demographic information and certain sources of income are not captured. For example, no current information return indicates whether a person is married or has a child. On the income side, information returns are particularly weak in capturing net income from self-employment.

The implications of this paper are potentially important to several income-based metrics and policies. Most importantly, the measurement of the United States income distribution may likely need to be adjusted as this research indicates low-income persons have much more income than previous survey-based research suggested. This could have implications for both the measurement of the poverty level and for measures of median income. Similarly, this new distribution of income could result in a re-examination of policies that rely on means testing such as school lunch programs and taxation of Social Security benefits. At the same

time, the large number of persons with wages who did not file an income tax return could indicate that far more people are eligible for income support programs, such as the Earned Income Tax Credit, than are currently receiving benefits.

Notes

- ¹ The four forms used by Sailer and Weber (1998), but excluded from this analysis, are the Forms 1099-S, 8596, W-2G, and 1098. This study uses five forms Sailer and Weber did not: Forms 1098-E, 1098-T, 1099-C, 1099-LTC, and 1099-Q.
- ² Since we are focusing on proportionality, it is worth noting that the size of individual age groups influence the relative size of other age groups to the whole. Thus, the factors causing young non-filers to be relatively fewer and older non-filers to be relatively greater magnify one another.
- ³ For the purposes of JCT's Individual Tax Model (but not this paper), "automatons" are imputed to our population of non-filers to fully represent the U.S. resident population. Automatons are simply placeholders and represent the non-filing population without information returns. At this stage, the only imputed information to an automaton are age and gender.
- ⁴ Total Income for non-dependents is Adjusted Gross Income before any above-the-line adjustments and without regard to the Parent's Elective Kiddy tax, the foreign earned income exclusion, or net operating loss carryovers. In addition, Total Income includes any untaxed Social Security and Railroad

Retirement Benefits, plus any deferred compensation. Certain unearned income items, which cannot be allocated between the primary taxpayer and spouse, are split 50-50. People with an overseas address are excluded.

References

- Centers for Disease Control and Prevention. National Center for Health Statistics. Table 1. Deaths, Age-adjusted Death Rates, and Life Expectancy at Birth, by Race and Sex; Maternal and Infant Deaths and Mortality Rates, by Race: United States, 2002 and 2003. Washington, D.C., 2003. <http://www.cdc.gov/>
- Cilke, James M. A Profile of Non-filers. Washington, D.C.: U.S. Department of the Treasury, 1998. Office of Tax Analysis Paper 78.
- Kilss, Beth and Fritz J. Scheuren The 1973 CPS-IRS-SSA Exact Match Study. *Social Security Bulletin* 41 (October 1978): 14-22.
- Sailer, Peter and Michael Weber. The IRS Population Count: An Update. Washington, D.C., Statistics of Income, Internal Revenue Service, 1998.
- Sailer, Peter, Michael Weber, & Ellen Yau. How Well Can IRS Count the Population? Washington, D.C., Statistics of Income, Internal Revenue Service, 1993.
- U.S. Census Bureau, Population Division. Population Estimates: Resident Populations. National Population Estimates for the 2000s. Washington, D.C., 2008. <http://www.census.gov/popest/national/asrh/2007-nat-res.html>

APPENDIX A

<i>Form</i>	<i>Components</i>
K-1	Business income, dividends, guaranteed payments, interest, long-term capital gains, ordinary income, other rental income, passive income, royalties, real estate, and short-term capital gains.
W-2	Wages, tips, FICA and Medicare wages and tax, advanced Earned Income Credit, deferred compensation, and dependent care.
5498	Individual Retirement Account (IRA), rollover IRA, Simplified Employee Pension (SEP), ROTH IRA, and recharacterized contributions; life insurance cost, fair market value of retirement accounts, and ROTH conversion amount.
1098-E	Student loan interest.
1098-T	Qualified tuition and related expense, scholarships or grants.
1099-B	Stocks and bonds, bartering, aggregate profit and loss, realized profit and loss, unrealized profit or loss.
1099-C	Amount of debt cancelled, property fair market value, interest forgiven.
1099-DIV	Capital gains, ordinary dividends, noncash liquid distribution, cash liquid distribution, non-taxable distribution, investment expense, ordinary dividend, foreign tax paid, and qualified dividends.
1099-G	Unemployment compensation, agricultural subsidies, prior year refund, and taxable grants.
1099-INT	Interest from savings, savings bonds, investment expense, and interest forfeiture.
1099-LTC	Gross benefits, accelerated benefits paid and long-term care date.
1099-MISC	Nonemployee compensation, medical payments, fishing income, rents, royalties, other income, substitute payments for dividends, excess golden parachute, crop insurance, and attorney fees.
1099-OID	Original issue discount (OID) and OID interest.
1099-PATR	Patronage dividends, non-patronage dividends, redemption amount, and retained allocations.
1099-Q	Gross distribution from a qualified tuition program and earnings on that gross distribution.
1099-R	Taxable amount of distributions from retirement investment vehicles, total employee contributions, and eligible capital gains.
1099-SSA	Total repayments, pension and annuity benefits, and workman's compensation offset.