

The Income Tax Position of Persons Not Filing Returns for Tax Year 2005

*Joshua Lawrence and Michael Udell, Ernst & Young LLP,
and Tiffany Young, Yale University¹*

This research focuses on the federal income tax position of persons resident in the United States who do not file an income tax return. For Tax Year 2005, about 38.6 million persons resident in the U.S. did not file or appear on a filed income tax return. A lack of sufficient income is the main reason that persons do not file a federal income tax return. However, 11.8 million of these persons did not file, or appear on a filed tax return but had sufficient income that required them to file. A significant number of persons who did not file an income tax return should have filed and owed tax. Of the estimated \$21.1 billion of total income tax liability associated with the 38.6 million persons who did not file a return, \$13.8 billion was unpaid. Many persons who do not file a tax return still pay taxes, however, and many of these persons would be due a refund of some or all of the income taxes withheld had they filed. Among the 38.6 million persons who did not appear on a filed income tax return, we estimate that if they didn't have any income not already reported to the IRS, then \$3.8 billion of income tax that was withheld by a third party on their behalf could have been refunded had they filed a tax return. In addition, many persons who did not file an income tax return could have been eligible for refundable tax credits such as the earned income tax credit and the child tax credit. As a result of not filing, as much as \$4.9 billion of refundable credits was not claimed. Importantly, all of these estimates are based upon amounts of income reported by a third party on behalf of a person and do not account for any unreported income.²

The goal of this research was to understand the federal income tax position of persons not filing an income tax return. A secondary objective was to rely, to the greatest extent possible, exclusively upon information that was filed with IRS as part of tax administration. The lone source of data not meeting this requirement was family structure information (married household with and without children; single head of household with children; and single household) necessary to infer the size of the resident U.S. population and filing status. The analysis focuses on Tax Year 2005, but can be repeated for any year in which third-party tax return information is filed on behalf of individuals.

This research began with an estimate of the U.S. residential population for 2005 from the March 2006 Current Population Survey (CPS).³ This population was then compared with the residential population shown on filed income tax returns for Tax Year 2005. Approximately 38.6 million persons resident in the U.S. did not appear on a filed tax return for Tax Year 2005.⁴

An estimated 30.8 million out of the 38.6 million persons who did not appear on a filed income tax return had a third-party information return filed on their behalf. This was estimated by drawing a random sample of all information returns filed for all persons for Tax Year 2005 and removing those information documents associated with a filed income tax return. By adding these 30.8 million persons to the 256 million persons who filed and appeared on income tax returns for 2005 and were resident in the U.S., we estimate that approximately 97 percent of the 296.5 million persons resident in the U.S. during 2005 were represented in the federal income tax system. Of the 38.6 million persons who did not appear on a filed income tax return, the federal tax system had some information from third-party filed information returns on 30.8 million and no information through third-party filed information returns on another 7.8 million.

If all 38.6 million persons were to have appeared on income tax returns, we estimate that there would have been an additional 22.8 million tax returns with the following income and tax characteristics:

\$511.5	billion of total income from information documents
\$233.7	billion of total income subject to tax
\$21.1	billion of income tax liability
\$11.2	billion of tax payments made
\$13.8	billion of income tax due before either child or earned income tax credits
\$3.8	billion of refunds due before either child or earned income tax credits
\$4.7	billion of potential earned income credit
\$2.8	billion of potential child tax credit
\$13.0	billion of income tax due after both child and earned income tax credits
\$8.7	billion of refunds due after both child and earned income tax credits.

The Data

The data set created for this project consists of a random sample of all information returns filed for Tax Year 2005 for persons regardless of whether a tax return was filed. Some of the approximately 256 million persons reported on a filed tax return had no information returns filed on their behalf. For example, many children do not have information returns but most appear on a filed tax return, whether timely-filed or late-filed. No-return persons (i.e., those who do not appear on a timely-filed or late-filed tax return) who had an information report filed on their behalf totaled 30.8 million, of which 30.4 million contained information about income and an additional 0.4 million contained information about attendance at an educational institution.

The 30.4 million no-return persons with income on information returns include:

- 16 million with social security benefits
- 10 million with wage income
- 8 million with interest income
- 6 million with pension, annuity, or IRA income

Together, the 256.0 million persons who appeared on filed tax returns and the 30.8 million no-return persons who had information reports provided by a third party indicate that the income tax system for Tax Year 2005 accounted for 286.8 million persons out of a total U.S. population of approximately 296.5 million (i.e., 97% of the U.S. population).

The information documents for the research consist of a sample of all information documents filed for individuals for Tax Year 2005. The sample included all of the information documents filed on behalf of a person and was based upon a random sample of social security numbers. Included in the sample were 1,256,000 information documents of which 1,121,000 (89 percent) were associated with timely filed income tax returns; 52,000 (4 percent) were associated with late-filed income tax returns; and 83,000 (7 percent) were associated with persons who did not appear on a tax return. In all, 28 different information documents were sampled including all Forms W-2, all Forms 1099 and 1098, all Forms K-1, and all Forms 5498.

The information returns report amounts of wages, interest, dividends, pensions and annuities, social security benefits, gambling winnings and partnership, S-corporation, and trust distributions, as well as gross proceeds from sales of capital assets and certain real estate transactions. In addition, any amounts of income tax withheld at the source and any amounts of estimated taxes paid were also included in the analysis. Some of the forms don't show amounts of income but rather amounts of interest paid on a mortgage or attendance at an institution of higher learning. The most important source of income absent from the data set was net income from a business operated as a sole-proprietorship, since this is a self-reported amount and no information return was filed to document that income.⁵ Thus, to the extent that business income is not reported through third-party information reporting, this analysis will understate income.⁶ The last piece of information provided for the data set was the age and gender of the person for whom the information report was filed.

Constructing Tax Families for No-Return Persons

Concepts

Information documents are filed for individual persons but tax returns can be filed for a number of living and support arrangements, including families (joint filing and head-of-household filing) as well as single persons.⁷ No information about the living and support arrangements for persons who did not appear on a filed income tax return is available from third-party information returns. Yet this information is necessary to determine filing status and ultimately whether a no-return person should have filed a tax return.

To bridge this information gap, hypothetical tax returns were constructed by placing each of the 38.6 million no-return persons into a family structure using the extensive living arrangement information contained annually in the March Current Population Survey (CPS). The CPS does not identify the tax filing status of a person. However, it is useful as a national sample of households that contains extensive information about the relationships among persons residing in a household. For example, married couples are identified as well as their dependents residing in the household. This research exploits these relationships to identify the family structures that could result in the tax filing status as married filing jointly, head of household, or single.⁸

In using the CPS data, the analysis focuses on the person in each household who is the CPS survey's reference person and analogizes that person as the primary person on a filed tax return. In a few cases there were households with multiple families each of which could result in a possible tax return. In these instances multiple tax returns could result from within a single CPS household. The demographics from these multiple tax returns—ages and number of persons in each family—would be used to identify a person like a CPS reference person. This newly identified reference person would also form the basis of a family structure from which a tax filing unit would be created from the relationship information in the CPS. These constructed CPS reference persons, along with the official CPS reference persons and their family structures, were used to match against the primary persons identified in the information returns sample by the IRS.

Each family structure within a CPS household that could result in a tax filing unit was matched to the primary person and their family structure from filed tax returns to account for all 256 million resident persons who appeared on filed income tax returns for 2005.

All of the no-return persons who had information documents were also eligible to be matched as primary persons to the CPS reference persons.⁹

Implementation

The matching of CPS data to information return data used a predicted mean matching algorithm as follows: First, the entire CPS was sorted by age and income of persons, and then all of the persons in the entire information returns file (timely filers, late filers, and no-return persons) were sorted by age and income. This sort by age and income results in some no-return persons interleaved among timely and late filer persons as a result of their greater amounts of income. For example, a 36-year-old primary filer from a timely filed return might have had total income on information documents of \$10,000, while a 36-year-old no-return person had total income on information documents of \$50,000. After both files were sorted they were matched to one another by relative ranking order according to total income.¹⁰ As a result, the no-return person would be matched with a 36-year-old CPS reference person with greater CPS income than would the 36-year-old timely filer person who had \$10,000 of total income.

The IRS had provided the family sizes and ages of dependents for primary persons who filed a tax return and who were included in the sample of information returns. A primary person on a tax return in our sample (and their filing family size, ages of spouse and dependents, and their income) was matched to a similar CPS reference person (and their family size, age of spouse and dependents, and their income). The matched CPS reference person and their family members were then removed from the CPS file. The remaining CPS persons provide a demographic portrait of the no-return population of 38.6 million persons.

The predicted mean matching algorithm had also matched CPS reference persons to some no-return persons. The family structures and the ages of persons related to the CPS reference persons were then replicated from the inventory of no-return persons by drawing without replacement from the no-return persons who were not matched to CPS reference persons. These persons became the tax family members of the no-return person who was matched to the CPS reference person. The incomes of these persons from the information returns data became part of the total income of the tax return for the family. Tax families could result in married filing joint returns; head-of-household returns; or single filer returns. The CPS had 38.6 million persons while the no-return file accounted for 30.8 million persons, so 7.8 million additional persons were needed to complete the modeling of the CPS family structures. No information was created for the 7.8 million persons other than their ages.

One desirable feature of this algorithm is that not all of the CPS persons who had income were matched to persons who were either timely-filers or late-filers, if a no-return person had greater income than either of these types of filers. This was the case as many persons file tax returns when they do not have a filing requirement, and some persons do not file tax returns even when they do have a filing requirement. The approach used here allows for no-return persons to reach well up into the income distribution reported on the CPS. What is most important in this predicted mean matching application is that the relative incomes of persons from the CPS be matched with the relative income distribution from the tax return information. The absolute difference in income between a CPS family and a tax return constructed family does not have a role.¹¹ Total income for a return was calculated as the sum of all of the reported amounts on information returns for persons in the constructed family.¹² Total income was used to test whether there was a requirement to file a tax return.

The 38.6 million no-return persons were distributed to 22.8 million simulated tax returns as follows:

TABLE 1. Filing Status and Number of Simulated Returns of No-Return Persons, TY 2005

Filing Status	Married Filing Jointly	Single	Head of Household	Total
Number of Returns	7,265,506	13,663,951	1,856,469	22,785,926
Number of Persons:				
Total	19,939,576	13,663,951	5,029,472	38,633,000
21 or younger	5,274,781	—	2,693,728	7,968,509
14 or younger	3,956,874	—	1,879,345	5,836,219

Source: Ernst & Young LLP analysis of IRS data

NOTE: Details may not sum to totals due to rounding.

Constructing Tax Liability for No-Return Persons

Total income is but a first step in determining tax liability. Both adjustments to income and deductions are necessary in order to calculate liability. However, information returns generally report amounts of income received, not amounts paid that would be allowable as adjustments or deductions. One source of data within the IRS to model amounts of adjustments and deductions on tax returns is the Statistics of Income Division individual income tax file for Tax Year 2005. These data include information about income, adjustments and deductions for filed tax returns and were used to impute total amounts of adjustments and deductions on each return.¹³

Separate imputations were performed for adjustments and deductions for each of the filing statuses created as a consequence of placing no-return persons into family structures: married filing joint; head-of-household; and single. The SOI individual income tax file was first censored by removing tax returns with total income amounts greater than the largest no-return person tax return. This exclusion avoided out-of-scope information from influencing the imputations of adjustments and deductions. Since filed income tax returns include 1) returns that were required to be filed as well as 2) returns that were not required to be filed (due to insufficient income), there is some similarity between the filed tax returns and the potential returns of no-return persons, justifying them as a basis for imputing adjustment and deduction amounts.

Adjustments were imputed beginning with the frequency and dollar amounts of total adjustments from filed returns. Adjustments were calculated for each return type (married filing jointly, head-of-household, and single) and age of primary taxpayer and income range using the following two-step procedure. First a logit model was estimated as to whether there were adjustments or not, and second, conditional on the logit parameters, an amount of adjustments was estimated. Adjustment amounts were quite small for the no-return person potential tax returns.

Imputing itemized deductions followed a similar approach with a few exceptions. Unlike adjustment amounts that lack a third-party information return, there was/is information on state and local taxes withheld and on mortgage interest paid. Both are significant components of itemized deductions. Following in the same manner as adjustments, first a logit model on the choice to use the standard deduction or the itemized deduction was estimated using these data. Conditional on the logit parameters, the amount of total itemized deductions was imputed. Itemized deduction amounts were more frequent than adjustments, largely due to the presence in the no-return data of many home mortgage interest payments on Form 1098-T and payments of state and local income taxes on Form W-2.

With filing status, total income, adjustments, and deductions determined, tax liability was calculated. Offsetting tax liability were amounts of taxes withheld and estimated tax payments made (as provided by the IRS). The result was either a balance due or refund amount, or no balance.

The Tax Position of No-Return Persons

General Results Before Child Tax Credit and Earned Income Credit

Table 2 shows that of the 22.8 million simulated tax returns representing 38.6 million no-return persons, 5.2 million (23%) had a filing requirement, a tax liability of \$21.1 billion, refunds due of \$2.9 billion, and tax due of \$13.8 billion. Of the \$233.7 billion of total income potentially subject to tax, \$196.7 billion (84%) was on these returns. Total income potentially subject to tax excludes amounts of non-taxable income, consisting mostly of social security income and some pension and annuity income.

TABLE 2. Tax Position of Simulated Tax Returns of No-Return Persons Before Child Tax Credit and Earned Income Tax Credit by Filing Requirement for TY 2005

[Amounts in Millions of Dollars]

Tax Position	Number of Returns	Number of Persons	Total Income	Distribution of Total Income	Total Tax Liability	Total Refund	Total Balance Owed
Filing requirement	5,180,474	11,822,423	\$196,694	85%	\$21,148	-\$2,909	\$13,811
Balance due	2,489,302	4,898,088	\$127,778	55%	\$17,891	\$0	\$13,811
No balance nor refund	670,967	1,425,122	\$6,236	3%	\$1	\$0	\$0
Refund	2,020,205	5,499,213	\$62,680	27%	\$3,256	-\$2,909	\$0
No filing requirement	17,605,452	26,810,577	\$37,041	16%	\$0	-\$917	\$0
No balance nor refund	14,496,123	20,550,448	\$23,516	10%	\$0	\$0	\$0
Refund	3,109,329	6,260,129	\$13,525	6%	\$0	-\$917	\$0
Total	22,785,926	38,633,000	\$233,735	100%	\$21,148	-\$3,826	\$13,811

Source: EY analysis of IRS data.

Note: Details may not sum to totals due to rounding.

The remaining 17.6 million potential returns (77%) were not required to be filed for income tax reasons, and accounted for 26.8 million no-return persons (69%) but only \$37 billion (16%) of total income potentially subject to tax. These returns had no income tax liability, but would have been eligible for \$0.9 billion in refunds prior to calculating the earned income or child tax credits.

TABLE 3. Tax Position of Simulated Tax Returns for No-Return Persons Before Child Tax Credit and Earned Income Tax Credit by Total Income for TY 2005

[Amounts in Millions of Dollars]

Total Income	Number of Returns	Number of Persons	Total Income	Distribution of Total Income	Total Tax Liability	Total Refund	Total Balance Owed
Less than \$0	46,620	84,378	-\$525	0%	\$0	-\$15	\$0
\$0–\$1,000	10,246,223	14,156,661	\$1,420	1%	\$0	-\$163	\$0
\$1,000–\$5,000	4,682,284	7,438,069	\$12,388	5%	\$0	-\$241	\$0
\$5,000–\$10,000	2,578,042	4,393,008	\$18,460	8%	\$13	-\$342	\$5
\$10,000–\$20,000	2,069,468	4,078,717	\$29,548	13%	\$520	-\$626	\$282
\$20,000–\$30,000	1,062,587	2,431,272	\$26,064	11%	\$1,080	-\$615	\$525
\$30,000–\$40,000	686,449	1,776,836	\$23,747	10%	\$1,347	-\$478	\$583
\$40,000–\$50,000	383,297	1,038,386	\$17,106	7%	\$1,284	-\$265	\$563
\$50,000–\$100,000	798,718	2,485,362	\$54,210	23%	\$5,093	-\$807	\$2,145
Over \$100,000	232,237	750,310	\$51,316	22%	\$11,811	-\$273	\$9,709
Total	22,785,926	38,633,000	\$233,735	100%	\$21,148	-\$3,826	\$13,811

Source: EY analysis of IRS data.

NOTE: Details may not sum to totals due to rounding

As shown on Table 3, the simulated tax returns for no-return persons can be sorted into three bins according to income and tax liability.

The lowest income bin, spanning up to \$10,000 of total income potentially subject to tax, accounts for 26.1 million (67%) of the no-return persons, with a total tax liability of \$13 million and a balance due of \$5 million. This same group of persons, however, is eligible for \$1.4 billion of refunds (36%).

The middle bin, spanning income between \$10,000 and \$50,000, accounts for 9.3 million (24%) of the no-return persons. These returns had an estimated total tax liability of \$3.7 billion and a balance due of \$1.7 billion but an almost equal amount of refunds at \$1.4 billion.

The 3.2 million persons (8%) associated with potential returns having total income potentially subject to tax in excess of \$50,000 account for \$16.9 billion (80%) of the \$21.1 billion in tax liability, \$11.9 billion of tax due, and \$1.1 billion in income tax refunds before the child tax credit. *Within this group are the largest 1,000 unweighted returns in the sample, shown on Table 4. These returns, weighting to 0.588 million in the population and accounting for only 1.9 million no-return persons (5%), had total income subject to tax of \$80.0 billion (34%), total liability of \$14.8 billion (61%) and the lion's share of tax due of \$11.0 billion (80%).*

Nevertheless, there are still sizeable refunds even for this narrowly defined group of high-income simulated returns, as nearly \$0.7 billion in refunds (17%) are to be found here. What do we know about these returns? Income on these returns is overwhelmingly from wage and salary compensation reported on the Form W-2, at \$61 billion (76%), followed by rental income, royalty income, and real estate income reported on Form K-1 at \$8.6 billion (11%), followed by business income of \$6.4 billion (8%) reported on Forms 1099-MISC and Form K-1. Only 4% of the income of the top 1,000 records in this data set is from taxable social security benefits (\$2.6 billion).

Social Security Beneficiaries

Many no-return persons are 62 or older and eligible for social security benefits. Another big story in these data is the tremendous amount of social security benefits that are not subject to income tax and not included in total income subject to tax on line 22. Almost 40% of no-return persons (15.139 million) are greater than 61 years old, but have potentially taxable income of only \$43.0 billion (18% of total) despite having social security income totaling \$151.3 billion. *Of this amount only \$2.6 billion, or less than 2%, is includable in total income subject to tax.* Total tax liability is \$2.1 billion, which is just 10% of the total tax liability of no-return persons (\$21.1 billion). Table 5 shows the income distribution of the elderly no-return persons:

TABLE 4. Total Income on Largest 1000 Simulated Tax Returns (Unweighted) of No-Return Persons by Total Income for TY 2005

[Amounts in Millions of Dollars]

Total Income	Number of Returns	Number of Persons	Total Income	Distribution of Total Income	Total Tax Liability	Total Refund	Total Balance Owed
Less than \$0	—	—	\$0	0%	\$0	\$0	\$0
\$0–\$1,000	—	—	\$0	0%	\$0	\$0	\$0
\$1,000–\$5,000	—	—	\$0	0%	\$0	\$0	\$0
\$5,000–\$10,000	—	—	\$0	0%	\$0	\$0	\$0
\$10,000–\$20,000	—	—	\$0	0%	\$0	\$0	\$0
\$20,000–\$30,000	—	—	\$0	0%	\$0	\$0	\$0
\$30,000–\$40,000	—	—	\$0	0%	\$0	\$0	\$0
\$40,000–\$50,000	—	—	\$0	0%	\$0	\$0	\$0
\$50,000–\$100,000	356,025	1,135,513	\$28,666	12%	\$2,946	-\$379	\$1,319
Over \$100,000	232,237	750,310	\$51,316	22%	\$11,811	-\$273	\$9,709
Total	588,262	1,885,823	\$79,982	34%	\$14,757	-\$652	\$11,028

TABLE 5. Total Income for Primary or Spouse Filers 62+ Years of Age on Simulated Tax Returns of No-Return Persons by Total Income for TY 2005

[Amounts in Millions of Dollars]

Total Income	Number of Returns	Number of Persons	Total Income	Distribution of Total Income	Total Tax Liability	Total Refund	Total Balance Owed
Less than \$0	19,865	37,369	-\$225	0%	\$0	\$0	\$0
\$0–\$1,000	5,803,448	7,373,401	\$742	0%	\$0	-\$61	\$0
\$1,000–\$5,000	2,693,236	3,877,368	\$7,211	3%	\$0	-\$79	\$0
\$5,000–\$10,000	1,482,875	2,237,739	\$10,472	4%	\$1	-\$87	\$1
\$10,000–\$20,000	653,687	1,039,777	\$8,872	4%	\$67	-\$179	\$51
\$20,000–\$30,000	122,977	223,012	\$2,977	1%	\$79	-\$59	\$41
\$30,000–\$40,000	64,373	136,550	\$2,200	1%	\$97	-\$60	\$36
\$40,000–\$50,000	29,334	55,959	\$1,323	1%	\$76	-\$36	\$27
\$50,000–\$100,000	64,197	123,010	\$4,177	2%	\$379	-\$57	\$185
Over \$100,000	19,025	35,095	\$5,288	2%	\$1,359	-\$89	\$1,241
Total	10,953,017	15,139,280	\$43,037	18%	\$2,058	-\$707	\$1,584

Source: EY analysis of IRS data.

NOTE: Details may not sum to totals due to rounding.

But for the 19,025 tax returns with income in excess of \$100,000, there would be only \$0.3 billion of tax liability (3%) associated with 15.1 million (39%) no-return persons at least 62 years old. Most of the refund amounts are generated from small W-2 wage withholdings that, when taken in consideration with the general exemption of social security income, are overwithheld.

Age Profile of Persons on Simulated Tax Returns

The age profile of the primary person on the simulated tax returns for no-return persons on Table 6 also shows a skewed distribution toward the elderly. Approximately 38% of no-return persons are 62 or older. The no-return elderly have a dearth of taxable income accounting for just 17% (\$40.2 billion) of the \$233.7 billion of taxable income. Of course, this amount excludes approximately \$150 billion of social security benefits that the

elderly received in 2005. In contrast, 83% (\$11.5 billion) of the total balance owed (\$13.8 billion) of no-return persons is associated with primary filers between the ages of 30 and 61, who also account for 71% of total income subject to income tax (\$165.6 billion). This is where the bulk of taxable income is, as can be seen on Table 6, both of wages and other sources, and is where the bulk of the tax liability and balance due amounts occur.

Table 6. Age of Primary Filer on Simulated Tax Returns of No-Return Persons for TY 2005
[Amounts in Millions of Dollars]

Age	Number of Returns	Number of Persons	Total Income Subject to Tax	Distribution of Total Income	Total Tax Liability	Total Refund	Total Balance Owed
Under Age 18	158,492	162,396	\$36	0%	\$0	\$0	\$0
Age 18-22	619,402	829,568	\$4,312	2%	\$211	-\$89	\$91
Age 23-29	2,224,276	3,924,616	\$23,560	10%	\$1,570	-\$415	\$969
Age 30-39	2,770,975	6,690,577	\$59,360	25%	\$7,264	-\$955	\$5,496
Age 40-49	3,179,851	7,124,998	\$61,831	26%	\$5,680	-\$982	\$3,216
Age 50-61	3,071,198	5,197,422	\$44,389	19%	\$4,785	-\$720	\$2,837
Age 62-79	6,267,311	8,875,243	\$24,868	11%	\$907	-\$446	\$518
Age 80+	4,494,420	5,828,178	\$15,378	7%	\$731	-\$219	\$685
Total	22,785,926	38,633,000	\$233,735	100%	\$21,148	-\$3,826	\$13,811

Source: EY analysis of IRS data.

Note: Details may not sum to totals due to rounding.

Returns with Wages

Potential returns with wage income account for 46% of all no-return persons (17.8 million), 77% of total income potentially subject to tax (\$178.8 billion), 68% of balance due amounts (\$9.4 billion) and 85% of refunds (\$3.2 billion) as shown on Table 7. More revealing, however, are the 0.176 million returns (less than 1.0% of simulated returns) with 605,601 persons (less than 1.6% of no-return persons) and total income subject to tax greater than \$100,000. These returns include \$36.9 billion of total income (16% of the total) and account for 45% of the balance due amount (\$6.2 billion). *Less than 2% of no-return persons account for almost half of all balance due amounts, and all of the simulated returns for these persons include wage income.*

Earned Income Credit

A surprisingly large number of no-return persons, almost 12.3 million, are associated with simulated returns potentially eligible for the earned income tax credit. The information documents alone are not sufficient to establish eligibility for the earned income tax credit, and this analysis relies mainly upon income (as established with information documents) and family structure (as identified with the CPS data) to estimate an upper bound of eligibility and credit amounts. The 12.3 million represent 32% of no-return persons. These persons account for only 19% of total income subject to tax (\$45.1 billion), 4% of tax liability (\$0.8 billion) and 3% of tax due (\$0.4 billion), but fully 34% of income tax refunds (\$1.3 billion). *In addition, the amount of "unclaimed" earned income credit could have increased the total amount of earned income credit provided in TY 2005 by an additional \$4.7 billion (14%) from \$34.5 billion actually claimed to \$39.2 billion.*

TABLE 7. Simulated Tax Returns of No-Return Persons with Wages by Total Income for TY 2005

[Amounts in Millions of Dollars]

Total Income	Number of Returns	Number of Persons	Total Income Subject to Tax*	Distribution of Total Income	Total Tax Liability	Total Refund	Total Balance Owed
Less than \$0	9,841	21,115	-\$135	0%	\$0	-\$15	\$0
\$0–\$1,000	1,023,440	1,738,294	\$454	0%	\$0	-\$17	\$0
\$1,000–\$5,000	1,668,576	3,135,996	\$4,492	2%	\$0	-\$152	\$0
\$5,000–\$10,000	1,098,868	2,166,717	\$8,109	3%	\$11	-\$260	\$4
\$10,000–\$20,000	1,442,981	3,051,646	\$21,106	9%	\$432	-\$495	\$203
\$20,000–\$30,000	947,791	2,231,718	\$23,243	10%	\$969	-\$564	\$431
\$30,000–\$40,000	620,856	1,650,094	\$21,475	9%	\$1,213	-\$462	\$468
\$40,000–\$50,000	347,518	955,335	\$15,485	7%	\$1,177	-\$248	\$471
\$50,000–\$100,000	704,904	2,268,897	\$47,668	20%	\$4,431	-\$758	\$1,577
Over \$100,000	176,845	605,601	\$36,913	16%	\$8,296	-\$266	\$6,227
Total	8,041,620	17,825,414	\$178,809	77%	\$16,530	-\$3,236	\$9,380

Source: EY analysis of IRS data.

NOTE: Details may not sum to totals due to rounding.

* Income Subject to Tax is the amount shown on line 22 of the Form 1040.

TABLE 8. Simulated Tax Returns of No-Return Persons with Earned Income Credit But Before Child Tax Credit by Total Income

[Amounts in Millions of Dollars]

Total Income	Number of Returns	Number of Persons	Total Income Subject to Tax*	Total Tax Liability	Total Refund Before EIC	Balance Owed Before EIC	Total EIC	Balance Due After EIC
Less than \$0	6,999	15,748	-\$61	\$0	-\$1	\$0	-\$3	\$0
\$0–\$1,000	1,287,474	2,103,774	\$518	\$0	-\$20	\$0	-\$394	\$0
\$1,000–\$5,000	1,909,177	3,346,594	\$5,070	\$0	-\$147	\$0	-\$768	\$0
\$5,000–\$10,000	1,156,124	2,141,339	\$8,478	\$11	-\$244	\$4	-\$964	\$0
\$10,000–\$20,000	785,156	2,158,467	\$10,638	\$60	-\$279	\$30	-\$1,591	\$0
\$20,000–\$30,000	381,059	1,390,943	\$9,353	\$121	-\$336	\$41	-\$724	\$0
\$30,000–\$40,000	175,815	765,966	\$5,913	\$133	-\$197	\$40	-\$140	\$0
\$40,000–\$50,000	25,209	115,201	\$1,113	\$44	-\$33	\$12	-\$35	\$0
\$50,000–\$100,000	39,199	185,242	\$2,624	\$179	-\$54	\$60	-\$48	\$13
Over \$100,000	7,283	32,248	\$1,413	\$294	-\$5	\$234	-\$8	\$226
Total	5,773,495	12,255,521	\$45,059	\$842	-\$1,315	\$420	-\$4,674	\$239

Source: EY analysis of IRS data.

NOTE: Details may not sum to totals due to rounding.

* Income Subject to Tax is the amount shown on line 22 of the Form 1040.

Focusing only on returns with total income less than \$10,000, \$11 million of the \$13 million of total income tax liability on all simulated returns with income below this amount is on earned income tax credit returns. The income tax refunds on these same simulated returns prior to the earned income credit would have been an estimated \$0.4 billion and almost entirely from wage withholding. The earned income tax credit added an additional \$2.1 billion of potential refunds. These returns are not eligible for a refundable child tax credit (because adjusted gross income will not be greater than \$11,000) and so the earned income credit would be the primary source of a refund through the tax system for these persons.

The earned income credit reduces the balance due amount of \$420 million by 43% to \$239 million on a total tax liability of \$842 million and increases refunds from \$1.3 billion to almost \$6.0 billion. *The net effect on the tax position of all no-return persons as a result of the earned income tax credit is to reduce the balance due amount from \$13.8 billion to \$13.6 billion and increase refunds from \$3.8 billion to \$7.5 billion.*

Child Tax Credit

Only 7 percent of the simulated returns, with nearly 7 million persons (18%) could be eligible for the child tax credit. As with the earned income tax credit, there is not sufficient information to determine eligibility for the credit solely from the information returns. Assuming that every primary taxpayer on a simulated return was the primary source of support for a child claimed, Table 9 summarizes the potential universe of no-return persons with potential child tax credit eligibility.

TABLE 9. Simulated Tax Returns of No-Return Persons with Child Tax Credit But Before Earned Income Tax Credit by Total Income

[Amounts in Millions of Dollars]

Total Income	Number of Returns	Number of Persons	Total Income Subject to Tax	Total Tax Liability	Total Refund Before Child Tax Credit	Balance Owed Before Child Tax Credit	Total Child Tax Credit	Refundable Child Tax Credit	Balance Owed After Child Tax Credit	Total Refund After Child Tax Credit
Less than \$0	842	3,367	-\$15	\$0	-\$2	\$0	-\$2	-\$2	\$0	-\$3
\$0-\$1,000	—	—	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0
\$1,000-\$5,000	—	—	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0
\$5,000-\$10,000	—	—	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0
\$10,000-\$20,000	365,860	1,367,659	\$5,676	\$7	-\$161	\$3	-\$229	-\$222	-\$3	-\$383
\$20,000-\$30,000	351,305	1,369,285	\$8,614	\$85	-\$322	\$29	-\$535	-\$454	-\$52	-\$776
\$30,000-\$40,000	270,559	1,125,490	\$9,384	\$260	-\$281	\$87	-\$540	-\$323	-\$129	-\$604
\$40,000-\$50,000	164,466	713,231	\$7,333	\$333	-\$166	\$123	-\$357	-\$122	-\$112	-\$288
\$50,000-\$100,000	415,354	1,818,572	\$28,512	\$2,148	-\$534	\$803	-\$895	-\$86	-\$6	-\$620
Over \$100,000	125,128	558,572	\$27,975	\$6,265	-\$119	\$5,229	-\$274	-\$2	\$4,957	-\$121
Total	1,693,513	6,956,084	\$87,481	\$9,098	-\$1,585	\$6,275	-\$2,831	-\$1,212	\$4,655	-\$2,796

Source: EY analysis of IRS data.

NOTE: Details may not sum to totals due to rounding.

The child tax credit reaches higher incomes than the earned income credit (which phases out at just above \$37,000 of income in 2005). The 13.8 million persons on simulated returns eligible for the child tax credit are associated with \$9.1 billion of income tax liability, as compared with 12.3 million persons with \$0.8 billion of income tax liability on simulated returns potentially eligible for the earned income tax credit. The \$2.8 billion of child tax credits on these returns reduces the balance due amount by 25% from \$6.3 billion to \$4.7 billion and increases refunds by 76% from \$1.6 billion to \$2.8 billion.

The net effect on the tax position of all no-return persons as a result of the child tax credit and the earned income tax credit is a reduction in the balance due amount from \$13.8 billion to \$13.6 billion to \$13.0 billion, and the increase in potential refunds from \$3.8 billion to \$7.5 billion to \$8.7 billion.

Returns with Both Earned Income Credit and Child Tax Credit

One-half of the simulated returns potentially eligible for the child tax credit would also be eligible for the earned income credit. These 767,164 returns had one-fifth of the income shown on all child tax credit returns (or \$17.5 billion of \$87.5 billion), and one-fiftieth (or 2%) of tax liability (\$0.2 billion of \$9.1 billion). However, these same returns account for 37% (\$3.2 billion) of total refunds of \$8.7 billion.

TABLE 10. Simulated Tax Returns of No-Return Persons with Child Tax Credit and Earned Income Tax Credit by Total Income

[Amounts in Millions of Dollars]

Total Income	Number of Returns	Number of Persons	Total Income Subject to Tax	Total Tax Liability	Total Refund Before CTC and EIC	Balance Owed Before CTC and EIC	Total Child Tax Credit	Refundable Child Tax Credit	Total EIC	Balance Owed After CTC and EIC	Total Refund After CTC and EIC
Less than \$0	—	—	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0
\$0–\$1,000	—	—	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0
\$1,000–\$5,000	—	—	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0
\$5,000–\$10,000	—	—	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0
\$10,000–\$20,000	310,853	1,175,396	\$4,821	\$5	-\$139	\$3	-\$195	-\$190	-\$1,067	\$0	-\$1,396
\$20,000–\$30,000	296,610	1,158,502	\$7,291	\$73	-\$273	\$25	-\$458	-\$389	-\$617	\$0	-\$1,279
\$30,000–\$40,000	158,626	698,709	\$5,329	\$116	-\$180	\$32	-\$341	-\$234	-\$128	\$0	-\$543
\$40,000–\$50,000	1,076	4,829	\$45	\$1	-\$2	\$0	-\$2	-\$2	0	\$0	-\$4
\$50,000–\$100,000	—	—	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0
Over \$100,000	—	—	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0
Total	767,164	3,037,436	\$17,487	\$196	-\$594	\$60	-\$997	-\$815	-\$1,812	\$0	-\$3,222

Source: EY analysis of IRS data.

NOTE: Details may not sum to totals due to rounding.

Conclusion

Of the 38.6 million persons who did not appear on a filed income tax return (no-return persons) for Tax Year 2005, 30.8 million had at least one third-party information document filed on their behalf. This research created tax returns for no-return persons relying solely upon a sample of these information returns filed by employers, businesses, financial institutions, governments, real estate professionals, educational institutions, and trustees. Families were constructed for no-return persons using a predicted-mean-matching algorithm to map family structures (shown in the March Current Population Survey for the residential U.S.) to tax system information documents lacking that information. Tax returns were created relying upon these family structures. Tax liability was determined by imputing adjustments to income and itemized deductions, using information documents (for the home mortgage interest deduction found on the Form 1098, as well as state and local income taxes withheld), supplemented by actual levels of adjustments and itemized deductions found on filed tax returns.

The 38.6 million no-return persons resulted in 22.8 million tax filing units with \$233.7 billion of income potentially subject to tax. Of the 22.8 million tax filing units, 4.5 million had a tax liability amounting to \$21.1 billion; 2.5 million returns accounted for an estimated \$13.8 billion in tax due; and 5.1 million returns accounted for \$3.8 billion in unclaimed refunds. Using imperfect data to determine eligibility of the child tax credit and the earned income tax credit, an upper bound estimate of the impact of both of these credits for no-return persons would have reduced the amount of tax due from \$13.8 billion to \$12.9 billion and increased the amount of unclaimed refunds from \$3.8 billion to \$8.7 billion.

This research should not be construed as providing an estimate of the tax gap associated with persons who did not file a tax return. First, most income from self-employment does not have third-party information reporting and is not included in the data used in this analysis. As a result, total income for no-return persons is likely understated.¹⁴ Second, no upward adjustment to reported income is used to account for unreported income in this analysis. The IRS, in preparing tax gap estimates, adjusts detected amounts of unreported income to account for undetected income. This analysis does not include this kind of adjustment. If it did, it is likely that nonfiling gap would be larger than suggested by this study—not just because of the tax on the addition income, but also because that income would likely reduce the amount of credits among the nonfilers. With these caveats, this analysis shows that substantial amounts of income tax are due by persons who do not file tax returns and that substantial amounts of refundable tax credits are likely unclaimed as a result of not filing a tax return.

Endnotes

- ¹ Josh Lawrence, Michael Udell, and Tiffany Young performed this research for the IRS Office of Research, Analysis and Statistics pursuant to contract TIRNO09-K-0053. Each of the authors worked for the Quantitative Economics and Statistics group (QUEST) at Ernst & Young LLP. Tiffany Young is at the Yale School of Management. Michael Udell is the corresponding author at Michael.udell@ey.com. Views in this paper are those of the authors and do not reflect those of Ernst & Young LLP or Yale University. This research is a continuation of the research program introduced in “Attaching the Left Tail: A New Profile of Income for Persons Who Do Not Appear on Federal Income Tax Returns,” by Jacob Mortenson, James Cilke, Michael Udell, and Jonathon Zytnick, in Proceedings of the 102nd Annual Conference on Taxation, National Tax Association.
- ² In addition, this analysis does not identify how much of unpaid taxes were, or would be collected through IRS enforcement actions.
- ³ The CPS is a random sample of households resident in the United States. Unlike tax returns, it does not include the residential population in institutions, such as prisons and college dormitories, and it does not include U.S. citizens resident overseas, so it does not cover the identical federal tax return population.
- ⁴ In arriving at this estimate, the research program distinguished between persons who filed an income tax return late—as late as the end of 2008—from persons who never filed a return. The estimates presented here are only for persons who failed to appear on a filed tax return for Tax Year 2005 by 2008. The great majority of so-called late filers, persons who appeared on a tax return for Tax Year 2005 after October 15 of 2006 (when the second extension for timely filing of a 2005 return had eclipsed) had filed by 2007.
- ⁵ While net income from a sole-proprietorship is not reported by a third-party source, some gross income of sole-proprietors is provided by a third party. Generally, payments made to a trade or business that exceed \$600 per year are required to be shown on a Form 1099-MISC. Specifically, payments to attorneys, payments for fish purchased directly from fisherman, and most payments for health care made through an insurance company are reported on a Form 1099 and were included in the data set.
- ⁶ The IRS, when estimating the tax gap, uses various factors to account for unreported income. None of these factors, or unreported income multipliers, was used in this analysis so the estimates presented here will not be directly comparable to an estimate of the tax gap associated with no-return persons.
- ⁷ This research did not create returns for the married filing separate filing status because the CPS lacks information that could be used to create that status.
- ⁸ The CPS does not include information about the extent of support that one person provided to another in the household. Support tests are an important part of the ability to claim a dependent exemption for tax purposes. For this analysis it was assumed that if a person had a family relationship with a person then they would be in that person’s family. The analysis also looked to the amount of income reported on the CPS for a person to determine whether they would have sufficient income on their own to be required to file a tax return. Having sufficient income and being older than 21 would prevent a person from becoming a dependent on another person’s tax return for this analysis.
- ⁹ IRS Office of Research staff provided filing status, primary person, family size and age, and total income information from filed income tax returns for the 1,121,000 information reports in the sample associated with timely filed 2005 tax returns and for the 52,000 information reports in the sample associated with late filed 2005 tax returns.
- ¹⁰ During this matching process sometimes a record from one file is split into pieces with identical information but different weights so that the record weight for each person from the two files is the same. One of the challenges in implementing predicted mean matching algorithms is to define groups of returns from the two files with equal or nearly equal total weights so that record splitting does not result in unused records. When records from one file are not matched to records from the second file in a predicted mean matching exercise, the marginal distributions from the first file are not preserved, which defeats the purpose of the algorithm.

-
- ¹¹ The nature of CPS non-response of income amounts is not investigated in this research. Future analysis of the relationship between CPS non-response of income amounts and IRS third-party information reported amounts of income would refine the accuracy of the predicted mean matching algorithm by reducing a key source of variability between the CPS and the IRS tax return information.
- ¹² The reporting rules for income associated with dependent filers were modeled from IRS publications. In some cases, small amounts of income associated with a dependent on a return would not be included on the parents' return. In addition, some information returns that only report gross proceeds, such as the Form 1099-B (for gross proceeds from the sale of securities) and the Form 1099-S (for certain gross proceeds from the sale of real estate) were adjusted to reflect a "net" amount of income. For the Form 1099-S this meant that no amounts of income were reported in total income due to the
- ¹³ This assumes that no-return persons are eligible to claim adjustments and deductions of the same magnitude as did similar people who filed returns.
- ¹⁴ The understatement may not be as serious as it suggests because some self-employment income is provided on third-party information reports and the gross amounts of these sources of income are included in total income, rather than the net amounts that would be included in self-employment income.