Earned Income Tax Credit Participation Rate for Tax Year 2005

Dean Plueger, Internal Revenue Service

The Earned Income Tax Credit (EITC) was created in 1975 to offset certain taxes for working individuals with low to moderate income with children. The credit has been refined and expanded through the years, and, in TY2005, the maximum credit was $4,400. The credit is claimed by filing a tax return with the IRS. It is a refundable credit, meaning taxpayers are eligible to obtain a refund for any portion of the credit that is not offset by a tax liability.

There has long been interest in measuring the level of participation in the EITC program by both IRS and other researchers.[1] As the credit has grown in relative value to eligible participants, and its potential to lift taxpayers out of poverty has been recognized, advocates for persons/families with low to moderate income have increased their interest in measuring EITC participation.

In 1999, the IRS formed the Stakeholder Partnerships, Education, and Communication (SPEC) function in IRS, whose "business model incorporates an indirect approach to community outreach, tax return preparation assistance, financial literacy, and asset building. Their model emphasizes collaboration with organizations that have shared service objectives and the ability to reach targeted special populations. The majority of the community-based partners and coalitions serve the low-income, limited English proficient (LEP), disabled, elderly, and Native American populations."[2]

SPEC promotes EITC participation by using aggregate filing data the IRS produces annually on EITC claimants to inform community-based organizations of recipients' tax return characteristics. Advocate organizations also need data on the potentially eligible nonclaimant population to help focus their outreach—which is not present in claimant data.

Most attempts to measure participation have been focused at the national level. Karl Scholz published an article in 1994 titled "The Earned

[1] Participation can be divided into taxpayer participation and dollar participation. The distinction is necessary because taxpayers who are eligible for higher credit amounts are more likely to participate than taxpayers who are eligible for a minimal amount of EITC. Therefore, the dollar participation rate will generally exceed the taxpayer participation rate.

[2] United States Department of the Treasury, Internal Revenue Service (2007), "The 2007 Taxpayer Assistance Blueprint Phase 2," p. 87.

Income Tax Credit: Participation, Compliance, and Antipoverty Effectiveness" describing his effort to estimate the EITC participation rate for TY 1990.[3] Scholz concludes that between 80 percent and 86 percent of eligible households participated in the program, and that 1.3 million to 2.0 million eligible taxpayers did not claim the credit. The nonparticipating households were estimated using data taken from the Survey of Income and Program Participation (SIPP) for survey participants who appeared EITC-eligible and did not file a return with the IRS. At the time of Scholz's study, the IRS calculated and refunded EITC benefits to those who did not claim the credit when they filed a return and appeared eligible (instead of sending a notice). The IRS has since ceased this practice. The current study improves on that method by using a fact of claimant indicator supplied by the IRS. Additionally, in TY 1990, there was no EITC benefit for childless workers. In TY 2005, childless workers could receive a benefit up to $399.

In 2001, the Government Accountability Office (GAO) estimated taxpayer EITC participation to be 75 percent and dollar participation to be 89 percent for TY 1999 by using a multiple data source method that did not involve linked records.[4]

The previous research did not provide estimates below the national level. Today, there is a clear demand for credible measures of taxpayer participation at lower geographic levels.

As there have been no regional estimates of participation, some users of EITC data have computed their own estimates of the potential unclaimed EITC in their communities. Improper and inconsistently applied assumptions and computational methods resulted in widely disparate estimates across the country. To address the situation, the IRS formally estimated unclaimed and overclaimed EITC at the Zip Code level by assuming that participation and compliance were uniform across the country and applying the TY 1999 GAO estimates using an appropriate computational method. The IRS distributed these estimates in the fall of 2003. Given the assumptions behind the numbers, W&I Research believes the computation likely produced very inaccurate results. Estimates using this methodology are no longer produced. The IRS committed to developing a better method of deriving more geographic-specific estimates.

In 2004, two similar, but different, approaches were pursued to estimate participation below the national level. The first approach, termed the

---

[3] Scholz, John Karl (1994), "The Earned Income Tax Credit: Participation, Compliance, and Antipoverty Effectiveness," *National Tax Journal*. 47: 1, pp. 63–87.

[4] U.S. General Accountability Office, Earned Income Tax Credit Participation, GAO–02–290R, December 14, 2001. See http://www.gao.gov/new.items/d02290r.pdf.

Regression method, combined data from the IRS Individual Return Transaction File (IRTF) and EITC Compliance Studies, and Census Bureau Current Population Survey Annual Social and Economic Supplement (CPS ASEC) to build a regression equation that could estimate State-level participation for years in which no compliance data were available.[5] The second approach, termed the Exact Match method, uses the IRS IRTF and CPS ASEC and links the data from the two data sources. The remainder of this report details the methodology and results of the Exact Match method. At present, neither approach has produced reliable estimates of local EITC participation. National participation estimates have been produced using the Exact Match methodology, which could be used with the American Community Survey to produce similar results but with greater geographic specificity. Future work will determine whether the American Community Survey (ACS) is a feasible replacement for the CPS ASEC. If feasible, the ACS would allow for local statistics on potentially eligible, nonparticipants.

Participation Rates

Desirable Features of a Participation Estimate

Three factors affect the efficacy of computed participation rates.

Currency—It is desirable to have estimates that reflect present participation rates. Changes in eligibility criteria, the impact of current and new enforcement and outreach activities, and economic conditions likely influence taxpayer behaviors and could produce changes in taxpayer participation over time.

Resolution—Resolution refers to the precision or fineness of geographic detail the data source provides and is directly related to the sampling frame of the data sources. The higher the resolution, the smaller the geographic area for which estimates may be made. While most studies have measured EITC participation only at the national level, getting to local estimates is extremely important for advocacy groups and policymakers to target outreach and education efforts.

Completeness—How 'tight' (accurate) is the methodology that produces the participation estimate? Do the data sources provide sufficient information needed to determine/estimate EITC eligibility or EITC fact-of-

[5] The regression method was an internal analysis completed by W&I Research in 2004 that was not externally published.

claim?[6] Completeness includes both completeness of the data collection and the accuracy of the responses provided by survey participants.

The Exact Match methodology provides relatively current estimates. While this study examined TY 2005, in the future, results could be produced within 1 year of the close of a tax year.

The dataset used for this project (CPS ASEC) does not provide extensive resolution. At present, taxpayer participation estimates have not been produced below the Census Bureau divisions.[7] The availability of variables in the CPS ASEC to completely model EITC eligibility is described in a later section of the paper.

Data Needed To Estimate Participation

Two primary data elements are needed to estimate participation:

- The number of taxpayers who were legally paid EITC (numerator).
- The total number of taxpayers potentially eligible to receive EITC (denominator), including nonfilers.

Once these two data elements are known, the participation rate can be computed by dividing the numerator by the denominator. The denominator is estimated by applying the EITC rules to data in each survey record. The numerator can be estimated from IRS compliance studies (which produce estimates of the number of legal claimants) or derived from special projects like this Exact Match project with the Census Bureau.

Number of Taxpayers Potentially Eligible To Receive EITC

The only source of data that enables a researcher to estimate the number of taxpayers eligible for EITC is the Census Bureau. No other organization collects data on U.S. residents at the level of detail and geography needed to estimate the number eligible for the entire U.S. The Census Bureau has three products capable of estimating the number of taxpayers eligible for EITC. To determine EITC eligibility from any of the following datasets, a "tax filing unit" is constructed by combining the income of married persons and determining the number of children for each tax unit. Once all relevant information is assembled into one record, the EITC eligibility rules

[6] EITC fact-of-claim is defined as a taxpayer/respondent who has been proven to have been paid EITC.

[7] A division is a subregion of a Census Bureau Region. For example, the Northeast Region is composed of two divisions (New England and Mid-Atlantic). See http://www.census.gov/geo/www/us_regdiv.pdf.

are applied to the record, and those records that meet the requirements are selected and weighted to provide national estimates of the number of respondents eligible to receive EITC (denominator).

Of the following four Census Bureau datasets available to estimate the number of taxpayers eligible for EITC, CPS ASEC currently provides the best information from which to estimate eligibility. It is released annually. The sampling frame allows for estimates to the four U.S. regions. CPS ASEC contains rich information relative to family structure and income to allow a generally accurate determination of EITC eligibility.

American Community Survey

Primary advantage—Large sample size with over 3 million U.S. addresses surveyed annually.

Primary disadvantage—Income and family relationship of respondents are not as detailed as the CPS ASEC dataset.

Survey of Income and Program Participation (SIPP)

Primary advantage—Panel survey that collects detailed income and tax data, which allow for eligibility and recipient status to be determined from the same data source.

Primary disadvantage—Lag in the date from when the survey data are collected and when they are released. Data are not provided annually.

CPS Annual Social and Economic Supplement (CPS ASEC)

Primary advantage—Provides a rich set of information relative to family structure and income, and is conducted annually.

Primary disadvantage—Sample size is not as large as American Community Survey, which restricts the ability to produce State-level estimates.

Future studies will explore using the American Community Survey as the sample size is 10 times larger than the CPS ASEC and allows for more geographic resolution.

Number of Taxpayers who are Legally Paid EITC

The number of taxpayers legally paid EITC may be estimated using three methods. All three methods must determine both eligibility and EITC fact-of-claim.[8]

[8] It should be noted that none of the methods will always record the taxpayer's true income and therefore will likely overstate the estimate of taxpayers legally paid EITC.

Survey of Income and Program Participation (SIPP)—A panel in this Census Bureau survey directly asks whether the respondent claimed EITC on a tax return. If the researcher determines the respondent is EITC-eligible from data gathered from the survey and if the respondent reports claiming EITC, the respondent is classified as a legal claimant. The Census Bureau recently redesigned this survey, and it will continue to collect information related to tax filing, including EITC status. However, there is substantial concern that, when the new SIPP is fielded, respondents will not answer the EITC question accurately because they may not be aware of their EITC status.

Match Census Bureau Data to IRS Administrative Records (Exact Match)—A Census Bureau dataset (Decennial Long Form, American Community Survey, SIPP, or CPS ASEC) is matched to IRS administrative records that contain EITC fact of filing. The survey data are modeled to determine who is potentially EITC-eligible, and the IRS data are used to determine who was actually paid EITC. Once the match is completed, the resulting set identifies eligible recipients. The number of eligible recipients is compared to the number modeled eligible to determine the participation rate.

IRS EITC Compliance Studies—IRS commissioned and publicly released EITC compliance studies in TY 1997, TY 1999, and TY 2001 (National Research Program).[9] The compliance studies were stratified samples that did not include late EITC claimants; analysts weight the sample to arrive at the number of nonlate filing taxpayers who made a legal claim. Alternatively, an analyst may compute the percentage of taxpayers who made a legal claim and multiply that percentage against the total number of claimants, including late filers, to arrive at counts of taxpayers with a legal claim.[10]

Prior to TY 2005, none of the three methods that directly estimate the number of eligible recipients was conducted on an annual basis. SIPP was a panel study with study life cycles of several years and had a significant lag time to data release. The Exact Match was viable only in years in which the IRS provides the Census Bureau with administrative records that contain

[9] The IRS meets its need for current compliance information through the National Research Program (NRP). In 2000, IRS established the NRP office as part of its efforts to develop and monitor strategic measures of taxpayer compliance. NRP provides a statistically valid representation of the compliance characteristics of taxpayers. The IRS is currently conducting another NRP study of individual return reporting compliance for TY 2006 to TY 2008. Preliminary, TY 2006 results are expected by 2010.

[10] This method assumes late claimants have the same compliance rate as timely claims.

EITC fact-of-filing.[11] An EITC compliance study will be a component of the TY 2006 to TY 2008 NRP studies.

An indirect method of estimating the number of eligibles was developed by W&I Research, in conjunction with National Headquarters Research, and is known as the Regression Method. This method incorporates data from CPS ASEC, EITC Compliance Studies, and IRS administrative records to develop regression models that estimate EITC participation. The models are applied to years in which no EITC Compliance Studies were commissioned.

Methodology

Computing Population Eligible for EITC from the CPS ASEC

Both methods (Exact Match and Regression) used by the IRS to determine EITC participation rates rely on CPS ASEC to provide estimates of the number of taxpayers eligible to receive EITC.

Current Population Survey Annual Social and Economic Supplement (CPS ASEC)

The Current Population Survey is an annual survey of approximately 78,000 households nationwide. The Annual Social and Economic Supplement, formerly known as the March Supplement to the CPS, is an expanded sample that collects detailed income information. The population represented is the civilian noninstitutional population living in the United States. Members of the Armed Forces living off post or with their families on post are included if at least one civilian adult lives in the household. Most of the data from the CPS ASEC are collected in March, with some data collected in February and April.

Because the dataset has variables relating to family composition, it is relatively straightforward to construct tax-filing entities from this file. For example, a family of four, composed of a married couple with two children ages 10 and 12, can easily be combined into one tax filing unit consisting of the combined incomes of the two adults with two dependents (or qualifying children).

[11] This match may be conducted annually, contingent on the IRS, Treasury, and Census ability to fund the project and continuing agreement among the three entities to conduct such studies.

IRS Modeling of the CPS ASEC Public Use File

The Wage and Investment (W&I) Research Division of IRS models EITC eligibility using the CPS ASEC public use file. Person records are compiled into filing units, Adjusted Gross Income (AGI) and EITC earned income are computed, and EITC qualifying children are tallied. The person and household identifiers along with the modeled filing information were transmitted to the Census Bureau and merged onto the internal file. The Census Bureau also models filing units as part of a tax calculator. The units and qualifying children modeled by the IRS differ slightly from the Census Bureau modeled units. In certain situations, the IRS model maximizes EITC eligibility, per IRS rules, whereas the Census Bureau model minimizes overall tax liabilities. For purposes of this study, the W&I Research set of eligible tax units was used. A summary of filing units transmitted to the Census Bureau is in Table 1.

Table 1. IRS W&I Modeled Filing Units (Weighted CPS ASEC 2006), Millions of Filers

	0 QC [1]	1 QC	2+ QC	Total
Single	3.94	0.00	0.00	3.94
Head of Household	0.67	1.75	3.89	6.31
Married Filing Jointly	0.00	4.47	4.33	8.80
Total	4.61	6.22	8.22	19.05

[1] QC=Qualifying Children.

Assumptions and Limitations

While the CPS ASEC provides a vast majority of the data needed to determine eligibility (or ineligibility), it does not provide information on all factors related to EITC eligibility. Following are the data limitations and eligibility assumptions that were used by W&I Research and Census Bureau analysts when implementing the algorithm to identify individuals/families eligible for EITC.

Qualifying Children

Under TY 2005 tax law, a qualifying child is any child who meets all of the following conditions:

- **Relationship Test**—Must be a son, daughter, adopted child, stepchild, eligible foster child, or a descendent of any of them (for example, a grandchild); or a brother, sister, half brother, half

sister, stepbrother, stepsister, or a descendent of any of them (for example, a niece or nephew).

- **Age Test**—Under age 19, or under age 24 and a full-time student, or permanently and totally disabled, regardless of age.
- **Residency Test**—Lived with the taxpayer in the U.S. for more than half the year.

Relationship and Age Tests

The CPS ASEC provides the age and relationship of all household members in relation to the householder. In cases when at least one parent resides with a child, it is possible to identify the parent using a parent pointer, but, when no parent is present, the exact relationship of a child to adults may not be known (unless the child is a direct descendent, such as a grandchild).[12] All that would be known when the child is not a direct descendent is whether the child is related or not related to the householder.

Residency Test

To be a qualifying child for EITC, the child must have resided in the household for more than 6 months. The CPS ASEC does not include information on how much of the year the children lived with a potential EITC recipient. All children are considered to have lived with the EITC-eligible individual in the U.S. for the required length of time. Therefore, no children are disqualified in the modeling due to a disqualifying residency outside of the EITC-eligible individual's home.

Adjusted Gross Income Tiebreaker (Qualifying Child of More than One Taxpayer)

Under the TY 2005 definition of a qualifying child, a child may be the qualifying child of more than one taxpayer (i.e., the child's parent and the child's grandparent, if they lived in the same home). In this situation, only one person may claim the child for EITC.

IRS Publication 596 provides an example that demonstrates that the taxpayers may decide how to allocate the qualifying children:

> "You and your three children live with your mother all year long. You are 25 years old. Your only income was $9,000 from a part-time job. Your mother's only income was $20,000 for her part-time job. Your

[12] The variable is "a-parent" and identifies the parent of the child. The 2007 CPS ASEC (TY 2006) added two variables that allow for the identification of both parents.

> children are the qualifying children of both you and your mother because they meet the relationship, age, and residency tests for both you and your mother. Only one of you can use each child to claim EIC. However, you and your mother can split the three qualifying children between you. For example, you can use one child to claim EIC, and your mother can use the other two."[13]

When determining the assignment of qualifying children, preference was given to the parent(s). However, if the parent(s) had no income or had more than two qualifying children, the (additional) qualifying children were reassigned within the family (to a grandparent, uncle, etc.) if the parent's tax position was not harmed (increased tax liability). In these situations, the parent would have no known economic incentive to block the other family member from claiming the child(ren) for EITC. Because family members may decide how to allocate the qualifying children, it is possible that family members may try to maximize the amount of the total legal credit the household will receive; however, in the syntax as currently written, if the parent had income and fewer than three qualifying children, the children were kept with the parent.

Citizenship (Resident Status)

The CPS ASEC provides the U.S. citizenship status of all respondents. However, if the respondent is not a U.S. citizen, CPS ASEC does not provide whether the person is a legal resident with a Social Security number eligible for employment (which is a requirement for EITC eligibility), or if he or she is an illegal resident. Noncitizens are required to reside in the U.S. for the entire tax year in order to be eligible for EITC. CPS ASEC does provide the location of residence of the respondent 1 year prior to the interview date. Noncitizens who resided outside the U.S. in the previous survey were disqualified for EITC in the model. The remaining group of nonresidents was included in the group of eligibles, if they passed the remaining EITC requirements (income, age, etc.).

Income

The Earned Income Tax Credit is allowed to individuals/families based on the amount of earned income that they receive during a tax year.[14] Additionally, individuals/families are disqualified for the credit if their investment

[13] TY 2005 publication, page 18, example 3.

[14] Earned income generally consists of wages, salaries, tips, net earnings from self-employment, and gross income received as a statutory employee.

incomes exceed the maximum for the applicable tax year ($2,700 in TY 2005). The amounts of earned income, investment income, and AGI are estimated from the amounts reported by the survey participants.[15] However, some limitations of the CPS ASEC data exist. They include:

- A person who files Form 2555 relating to the receipt of foreign-earned income is not eligible for EITC. Since CPS ASEC does not include information on the receipt of foreign income or the filing of Form 2555, no individuals were determined to be ineligible for EITC due to the receipt of foreign income. In TY 2005, there were 308,000 total tax returns filed with Form 2555, with 191,000 having an IRS-computed AGI of $37,000 or less.[16]
- Individuals/families with investment income exceeding a certain threshold in the applicable year are not eligible for EITC. Investment income includes taxable interest income, tax-exempt interest income, dividend income, capital gains, and net income from rents and royalties derived from real estate. The amount of capital gains reported by CPS ASEC is imputed by the Census Bureau via a match with Statistics of Income (SOI) data. The match is not an exact match; rather, it is based on a "categorical" match and may not be accurate at the individual level. Therefore, the CPS ASEC estimate of capital gains was not included as investment income, and no individuals were determined to be ineligible for EITC because they received capital gains that would have caused them to exceed the investment income limitation.[17]

GAO conducted an analysis to determine the impact the missing capital gains may have had in the determination of eligibility for their TY 1999 study that estimated participation rates. Below is an excerpt from their TY 1999 participation study regarding this issue:

> "The CPS does not contain all of the information needed to determine eligibility. Data such as capital gains and contributions to individual retirement accounts are not requested in the CPS survey. However, it is likely that the missing data have little effect on our participation

[15] It should be noted that the householder reports the income for all members of the household and may provide estimates for persons for whom the householder has less than perfect knowledge.

[16] IRTF housed on the Compliance Data Warehouse.

[17] The CPS individual is assigned a category based on age, income, location, etc. Then, persons from the SOI public use file are grouped into the same categories. Then, the persons are matched on categories, and the capital gains from the individual in the SOI public use file are assigned to the person in the CPS data file.

estimates. These types of income and deductions are notcommon for the lower-income people who may be eligible for the credit. To verify this, we examined the tax returns of households in the 1996 Statistics of IncomePublic Use File, the most recent data available. We found that no more than 3 percent of households that met the other income limits for credit eligibility had these types of income and deductions."

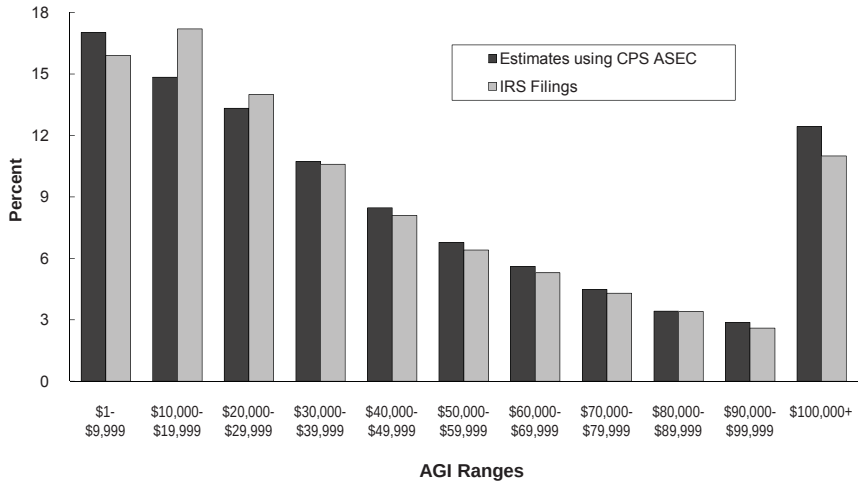
- Individuals whose AGI exceeds the limitations are not eligible for EITC. Several items used in determining AGI were not available in the CPS ASEC, including (but not limited to) taxable refunds, IRA and other retirement plan contributions, medical savings account deductions, moving expenses, self-employed health insurance deductions, penalties on early withdrawal of savings, and alimony paid. Income from trusts is also not available. Therefore, AGI will be understated for those individuals who had taxable refunds and capital gains and overstated for those individuals who had deductions and capital losses. Social Security income is included in the AGI calculation.
- Income computations are only as valid as the data provided by the respondent. Some respondents will intentionally or unintentionally provide incorrect information leading to incorrect estimates of income (earned income, investment income, and AGI). If the incorrect data are reasonably close to the true value, the number of taxpayers estimated to be eligible for EITC will not be significantly affected.[18] It is unknown if there are offsetting errors caused by some respondents underestimating their incomes and other respondents overestimating their incomes.

Finally, over the course of conducting the Exact Match, Census Bureau analysts discovered that about 18 percent of tax units estimated to be EITC-eligible had at least one income variable amount allocated. Amounts are assigned (imputed) when the respondent refuses or cannot provide the requested information. The amount imputed is based on the respondent's demographics and the income of others who reported their income in the same demographic category.

[18] However, the estimated amount of EITC the taxpaper is eligible to receive will have more error as the amount of EITC is directly related to income.

Figure 1 provides a comparison of modeled AGI and actual AGI for the population over 18 with positive AGI in TY 2005. The light bars represent AGI reported by taxpayers to the IRS, and the dark bars represent the AGI computed using CPS ASEC data.

Figure 1. Actual Tax Return and Estimated CPS ASEC AGI Comparison for TY 2005



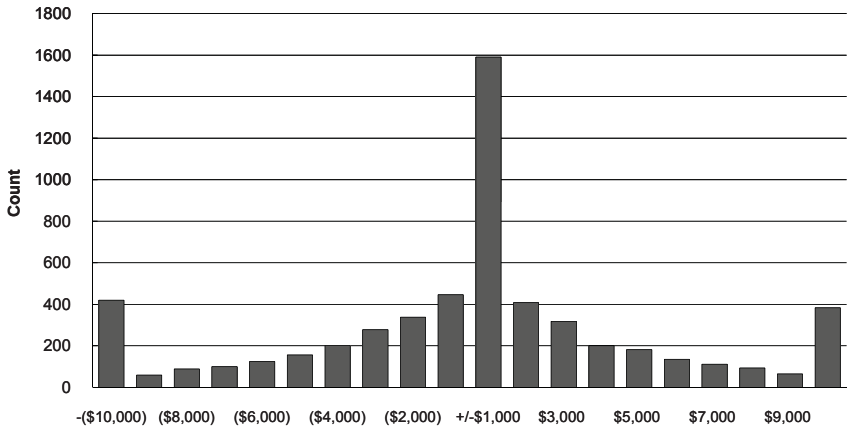
SOURCES: 2006 CPS ASEC and TY 2005 IRTF housed on CDW.

The difference in the $10,000 to $19,999 income category is potentially attributable to higher income taxpayers underreporting their incomes, therefore artificially raising the percentage of taxpayers in the $10,000 to $19,999 range. The difference may also be a result of tax units modeled in the CPS ASEC that do not file a tax return; and the W&I Research algorithm does not generate tax units for adult dependents (college students) as they are not eligible to receive EITC.

It should be noted that Figure 1 is not a result of a matched dataset. It was produced by creating two income distributions from the two datasets and combining them into one chart. Figure 2 is the result of matching the CPS ASEC set of taxpayers modeled eligible for and paid EITC to their tax return information. Figure 2 depicts the differences between the IRS AGI and CPS ASEC estimated AGI categorized into $1,000 increments. Differences of less than $1,000 (plus or minus) were grouped together. The group with the highest count had a difference in estimated and actual AGI of plus

or minus $1,000. For nearly every case where CPS ASEC AGI was overestimated, there is another case where it was underestimated, which explains why the aggregate AGI distributions shown in Figure 1 match as well as they do (the errors offset at the aggregate level). The counts of returns at the ends of the distribution with differences of at least $10,000 comprise 14 percent of all returns.

Figure 2. Difference between IRS AGI and CPS ASEC Estimated AGI



SOURCE: TY 2005 IRS-CPS ASEC Exact Match.

Filing Status

A taxpayer whose filing status is Married Filing Separate is not eligible to receive EITC. The filing status used for return preparation is not collected by the Census Bureau, but the survey does collect the person's current marital status (never married, married, divorced, and separated). Individuals who have no children, and who reported their marital status as separated, were assigned a filing status of Married Filing Separate and were not identified as eligible.

Taxpayers who were separated from their spouses for the final 6 months of the tax year and who have children are potentially eligible to use the Head of Household filing status. In the CPS ASEC data, individuals reporting children and a marital status of separated were examined to determine if their marital status was separated in the previous year (indicating they were likely separated the final 6 months of the year). If they reported their status as any status other than separated in the previous supplement, they were assigned a filing status of Married Filing Separate in the current

year.[19] For TY 2005, 4.9 percent of the tax units identified as eligible for EITC from CPS ASEC had a marital status of "separated."[20]

EITC is currently structured such that taxpayers using the filing status Married Filing Joint are able to receive a higher credit than those who file with other statuses when their AGIs or earned incomes are in the phaseout region of EITC eligibility. Families who had both spouses present in the household were assigned a marital status of Married Filing Joint and modeled to be eligible for the higher credit, when their incomes dictated.

CPS ASEC-IRS Exact Match Method of Estimating EITC Participation

Census Bureau Processing with the CPS ASEC Internal File

The CPS ASEC is an annual supplement to the basic monthly Current Population Survey. Approximately 10 percent of households in the sample fail to complete ASEC interviews. These cases are evaluated to determine whether sufficient information exists to impute ASEC responses from a similar case. Based on key demographic characteristics, all survey responses from another case are imputed to replace the missing data. The Census Bureau and the IRS agreed that these fully imputed cases were unsuitable for a record check analysis, as the data in these records would not be expected to match administrative record data.

Table 2: Incidence of Full Record Imputation (2006 CPS ASEC)

	Unweighted	Weighted	Percent of weighted
ASEC reported	189,112	264,170,000	89.9%
Fully imputed	19,450	29,664,000	10.1%
Total	208,562	293,834,000	100.0%

Source: Data Integration Division, U.S. Census Bureau

Person Identification Validation System

To enable the file linkage, the input files are processed through the Person Identification Validation System (PVS). The PVS compares identifying name, address, and date of birth due data from the CPS ASEC against a

[19] As a result of the sampling structure set up by CPS ASEC, approximately 50 percent of the households in this year's CPA ASEC dataset will be present in the previous year's CPS ASEC dataset.

[20] The predominant marital status was "never married" at 35.5 percent, followed by married (spouse present) at 32.9 percent.

reference file and assigns a unique identifier to records with verified data. Similarly, name, address, and SSN from the individual tax returns are compared to the reference file and assigned a unique identifier. In compliance with Census Bureau privacy policy, survey records lacking respondent consent for data linkage are not processed through PVS. Due to technical constraints, the PVS does not process records where the first name and last names are missing. The output file of the PVS process contains: all verified records; all nonverified records, including those where multiple matches were found; and all original records withheld from the PVS process due to linkage refusals or incomplete identifying data. Only validated and uniquely identified records are used in this study. A unique nine-digit protected identification key (PIK) is assigned to each validated record. The PIK is the linking key used in Census Bureau administrative record research projects; SSNs are not used.

The 2006 CPS ASEC survey marked an important change regarding the assignment of validated PIKs: It was the first survey year in which the SSN was not requested from respondents. In prior years, responses to the SSN question had fallen. This limited the number of persons who could enter the PVS process. New language obtaining consent from respondents—without asking for the SSN—permitted more cases to enter the PVS process and obtain PIKs.

Table 3: Refusal To Provide SSN or Linking Authorization (Unweighted Counts)

	2003 CPS ASEC	2004 CPS ASEC	2005 CPS ASEC	2006 CPS ASEC
Refused	36,793	49,026	50,846	259
% of Total	17.0%	23.0%	24.1%	0.1%
Total	216,424	213,241	210,648	208,562

Source: Data Integration Division, U.S. Census Bureau

Not all records receive PIK in this process. The primary components of the reference file are IRS and SSA Numident data. The Numident file only includes information on persons who have SSNs. The process fails to assign PIK to groups of persons, including undocumented residents, and persons with Individual Taxpayer Identification Numbers (ITINs). Persons with multiple matches also fail PVS and do not receive PIKs. A growing number of survey respondents fail to provide their first and last name data. These cases cannot enter the PVS process and lack PIK as well.

Table 4: PVS Results for 2006 CPS ASEC (Unweighted Counts)

No authority to link	259	0.1%
Missing first and last name	2,299	1.1%
Found in Geokey search	175,237	84.0%
Found in name search	9,221	4.4%
Multiple matches	468	0.2%
Not found	21,078	10.1%
Total	208,562	100.0%

Source: Data Integration Division, U.S. Census Bureau

The increased number of persons entering the 2006 CPS ASEC PVS process resulted in an increased number of persons with PIKs. The 2006 CPS ASEC increase in PIK affected adults more than children.

Table 5: Adult-Child Distribution of PIK Increase (Unweighted Counts)

	2004 CPS ASEC	2005 CPS ASEC	2006 CPS ASEC
Adults	103,777	100,076	139,958
Children	47,342	46,672	44,500
Total verified	151,119	146,748	184,458

Source: Data Integration Division, U.S. Census Burea

Table 6 provides a breakout of citizenship status of tax units modeled EITC-eligible and how many records were included in the final analysis.[21] Noncitizens were much less likely to be included in the final analysis because 29.5 percent were dropped as result of not being able to assign a PIK. This percentage is three times larger the CPS ASEC respondent universe (10.1 percent). This result raises questions about their initial eligibility determinations based on the modeling when their SSN status is unknown.

Table 6: Citizen Status of CPS ASEC Records (Unweighted Counts)

	Number Modeled Eligible	Number Remaining after Removal of Non-PIKed and Imputed Records	Percentage included in Analysis
Citizen	11,647	6,605	56.7%
Non-citizen	2,292	805	35.1%
Total	13,939	7,410	53.2%

Source: Data Integration Division, U.S. Census Bureau

[21] The CPS variable for citizenship status, PRCITSHIP, was used because the survey does not ask whether the respondent has an SSN. It is an imperfect proxy for having an SSN because some noncitizens can obtain an SSN, and some citizens may not have an SSN.

Data

Tax Returns from the Individual Returns Transactions File

The IRS annually provides the Census Bureau administrative records containing tax return information under 26 U.S.C. § 6103 (j)(1). The variables transmitted to the Census Bureau under this agreement are:

1. Name, address, and taxpayer identifying number of the taxpayer and spouse
2. Marital status
3. Number and type of exemptions (dependents)
4. Wages and salary income
5. Dividend income
6. Interest income
7. Gross rent and royalty income
8. Social Security income
9. Total of wages, interest, dividends, alimony, business income, pensions, rents, royalties, farm income, unemployment compensation, and Social Security benefits.
10. AGI
11. Indicator variables for Schedules A, C, D, E, F, and SE and Form 8814.

Other CPS ASEC tax model evaluation projects have used the data. This project uses the return level data on filing status, AGI, and number of children at home exemptions. The return level data are processed through the PVS, and records with validated information are assigned PIKs. ITINs on the file do not receive PIKs because those numbers are not present on the Numident file. Future refinements of our PVS process will address this issue, potentially reducing the number of non-PIK cases restricted from the analysis. The 1040 file is delivered for each filing year. Analysis proceeds when all 52 weeks are received at the Census Bureau.

The exact match occurs by linking records in the CPS ASEC to the 1040 file by PIK. This allows us to append tax return data to the survey record for the same individual. A key benefit to this analysis is the ability to evaluate 1040 filing and EITC-claiming behaviors of the EITC eligible

population, based on the CPS ASEC survey responses. For Tax Year 2005, 79,107 individual income tax returns matched to the nearly 140,000 CPS ASEC adults with PIKs. We are confident that the PVS produces high-quality PIKs on the survey data. The IRS data have very high-quality identifying information, which also results in high quality PIKs. Therefore, joining the files by PIK provides data from both agencies on the same subset of the population. The records that did not receive PIKs were not investigated in this study.

This transmission does not contain late-filed returns (for example, a TY 2005 return filed in Calendar Year 2007 would not be included in the normal Form 1040 delivery).

EITC Returns Extract

The data in the Form 1040 transmission allow fact of tax return filing to be determined for CPS ASEC respondents, but they do not identify which filers were paid EITC. In order to identify who was paid EITC, the IRS negotiated a contract under 26 U.S.C. § 6103 (n) with the Census Bureau. Under the contract, the IRS agreed to transmit four additional EITC variables for TY 2005 to the Census Bureau: earned income amount, number of EITC qualifying children, taxpayer-reported EITC amount, and IRS-computed EITC amount.

Taxpayers claiming EITC on their TY 2005 tax returns or subsequent amendments (including taxpayers paid EITC because of a CP-09/27 eligibility notice), through the end of Calendar Year 2007, were included in the data transmission to the Census Bureau. There were 23,296,704 records meeting these criteria. Although some TY 2005 EITC claims arrived after the cutoff date, this transmission accounted for 99 percent of EITC claims.[22]

This second set of IRS data was processed through PVS to assign PIK to enable data linkage, resulting in 14,081 returns matching PIK in the CPS ASEC. Input files from the IRS are processed in the Data Integration Division and protected per instructions in our Interagency Agreement and IRS Publication 1075.

The matched sample of EITC-eligible persons from the CPS ASEC data and IRS administrative data indicating who received EITC enable production of EITC participation estimates using our Exact Match methodology.[23]

---

[22] As of October 15, 2008, there were 23,465,092 TY 2005 returns filed meeting the condition.

[23] The accuracy of the estimates is discussed in the limitations section.

ITIN Extract

To investigate how missing ITINs could impact modeling efforts and the participation rate estimate, IRS transmitted an extract of 3,000,000 ITIN filers.[24] These administrative record cases were first linked to CPS ASEC persons on name and address, then on name alone. We wanted to check whether the survey data on citizenship and migration permitted accurate modeling of EITC eligibility. Any return with an ITIN, whether the primary or secondary filer, is ineligible for EITC. The name match determined that 565 persons were common between the files.[25] None of those persons had been modeled EITC-eligible.[26] While more research is needed on this subset, this preliminary finding indicates that the survey questions on citizenship may permit adequate modeling.

The original set of CPS ASEC tax units identified as EITC-eligible contains an unknown number of tax units that are not eligible for EITC—U.S. residents without an SSN valid for employment. The error of not being able to exclude these residents will upwardly bias the number of taxpayers eligible for EITC and the number of nonclaimants. The inclusion of additional ITIN data in future years may help reduce this known but unquantified bias.

Combined 1040-EITC Data

The combined 1040-EITC file was expanded from return level to person level. The combined file contained 41,824 single and head of household returns. Of the 79,112 returns, 37,288 were married filers. Both spouses received PIK in 33,083 of these returns, leaving 4,205 with only one spouse receiving PIK. The other spouses may have failed validation for a number of reasons: they may have lacked sufficient name or date of birth information, they may not have had an SSN, or they may have had multiple matches. After expanding the married returns to person records with PIK, the IRS data contained 112,195 records.

---

[24] ITINs used in the primary or secondary position on TY 2005 tax returns and their associated spouses, if present.

[25] The limited number of ITIN taxpapers matching the CPS ASEC database may be a result of limited information available to identify a CPS respondent. W&I Research is working to provide additional information for future matches to provide better identifying information.

[26] The vast majority of the validated ITIN population had a citizenship classification as foreign born, noncitizen (89 percent). Most taxpayers with an ITIN that were validated did not have imputed earnings (90 percent) and did not have self-employment income (97 percent); most were between the ages of 25 and 44 (55 percent); only three taxpayers were over the age of 55. Most taxpayers with an ITIN that were PIKed did not have a high school education (55 percent).

The 1040 and EITC data files containing PIKs present in the CPS ASEC were combined by survey person identifiers. The analytical files do not contain any personally identifiable information: no SSNs, no names, no addresses, no dates of birth, and no PIK, only survey person identifiers.

Allocated Earnings

CPS ASEC respondents are asked about their labor force participation and earnings. Persons who report being employed are asked for their hours and earnings. 10 percent of survey respondents do not report an earned income amount. Using other variables, earnings are allocated from another respondent with similar characteristics to the person with missing data. This hot-deck imputation produces $1.2 trillion in earned income, comprising 20.4 percent of total earnings in the survey. The income allocation in the CPS ASEC is not State-specific; a donor record is not necessarily from the same State or region as the recipient record. The allocated values are sufficient for analyses at the national level when viewing the entire earnings distribution. Allocated values are not designed for use at the person level. As a result, the allocated amounts often differ substantially from income reported to the IRS for the same person. The quality of allocated income amounts will be evaluated in the future. At this time, cases with allocated earnings were removed from the analysis.

Reweighting

The Census Bureau computes person weights for the CPS ASEC file that sum to the population controls for the civilian, noninstitutionalized population of the United States. By removing nearly one-quarter of the initial person records—those lacking PIKs, those with fully imputed data, and those with allocated earnings—the weights no longer aggregate to the population count. The removed cases are missing income data that are essential for modeling EITC eligibility, regardless of their reasons for being missing. As in other administrative record research projects, we assume that the data are missing at random and inflate the person weights on the remaining cases to reflect the population count. The procedure is similar to the nonresponse weighting adjustment the Census Bureau and other survey researchers use. Adjustment factors are calculated for the following grouped variables: age, marital status, race, and Hispanic origin. Some of the groups created by crossing all of these variables are very small. Cells are collapsed to bring the count (in each cell after the collapsing procedure) to at least 50 for the:

1) under age 24 and married; 2) age 65 and older, not married, and Hispanic; and 3) age 65 and older, married, and Hispanic. After calculating the adjustment factors for the groups, the appropriate factor is applied to each sampled person's survey weight, which the Census Bureau calculates for each person based on many characteristics. Data used to develop the adjustment factors are shown in Table 7.

Table 7: Reweighting Factors, by Partition (2006 CPS ASEC)

Age	Married	Hispanic	Race	Count	No PIK	With PIK	% w PIK
<24	N	N	White	46465	8677	37788	81.3%
<24	N	N	Black	8628	2285	6343	73.5%
<24	N	N	Aian [1]	943	265	678	71.9%
<24	N	N	A/Nhopi [2]	2851	934	1917	67.2%
<24	N	N	Other	2715	478	2237	82.4%
<24	N	Y	White	14018	4269	9749	69.5%
<24	N	Y	Black	444	137	307	69.1%
<24	N	Y	Aian	232	69	163	70.3%
<24	N	Y	A/Nhopi	108	28	80	74.1%
<24	N	Y	Other	507	129	378	74.6%
<24	Y	N	White	930	253	677	72.8%
<24	Y	N	Black	82	31	51	62.2%
<24	Y	N	All Other	78	30	48	61.5%
<24	Y	Y	All	586	354	232	39.6%
<24	N	N	White	20999	6139	14860	70.8%
24 to 64	N	N	Black	6536	2248	4288	65.6%
24 to 64	N	N	Aian	643	207	436	67.8%
24 to 64	N	N	A/Nhopi	1647	729	918	55.7%
24 to 64	N	N	Other	911	249	662	72.7%
24 to 64	N	Y	White	5414	2433	2981	55.1%
24 to 64	N	Y	Black	233	87	146	62.7%
24 to 64	N	Y	Aain	108	56	52	48.1%
24 to 64	N	Y	A/Nhopi	57	27	30	52.6%
24 to 64	N	Y	Other	146	59	87	59.6%
24 to 64	Y	N	White	51670	13067	38603	74.7%
24 to 64	Y	N	Black	4875	1730	3145	64.5%
24 to 64	Y	N	Aian	549	172	377	68.7%
24 to 64	Y	N	A/Nhopi	3963	1475	2488	62.8%
24 to 64	Y	N	Other	1068	261	807	75.6%

Footnotes at end of table.

Table 7. Reweighting Factors, by Partition (2006 CPS ASEC)—Continued

Age	Married	Hispanic	Race	Count	No PIK	With PIK	% w PIK
24 to 64	Y	Y	White	10089	4194	5895	58.4%
24 to 64	Y	Y	Black	213	99	114	53.5%
24 to 64	Y	Y	Aian	144	62	82	56.9%
24 to 64	Y	Y	A/Nhopi	78	34	44	56.4%
24 to 64	Y	Y	Other	219	79	140	63.9%
65+	N	N	White	6259	1050	5209	83.2%
65+	N	N	Black	1393	299	1094	78.5%
65+	N	N	Aian	101	20	81	80.2%
65+	N	N	A/Nhopi	375	95	280	74.7%
65+	N	N	Other	143	17	126	88.1%
65+	N	Y	All	805	169	636	79.0%
65+	Y	N	White	8679	1573	7106	81.9%
65+	Y	N	Black	877	171	706	80.5%
65+	Y	N	Aian	54	12	42	77.8%
65+	Y	N	A/Nhopi	584	185	399	68.3%
65+	Y	N	Other	187	29	158	84.5%
65+	Y	Y	White	900	191	709	78.8%
65+	Y	Y	NonWhite	56	11	45	80.4%

[1] American Indian or Alaska Native
[2] Asian/Native Hawaiian or Other Pacific Islander
Source: Data Integration Division, U.S. Census Bureau

Record Linkage of IRS and Survey Data

The reweighted CPS ASEC file was matched to the IRS 1040 and EITC files by PIK, retaining survey person and household identifiers. This file contained the modeled tax filing units to be used in determining eligibility. As the modeled return information was assigned to the presumed tax filer, the file could now be collapsed back to the return level. The file was unduplicated at this point. A preliminary review indicated that many duplicate returns were due to second returns being filed to correct filing status and income amounts.

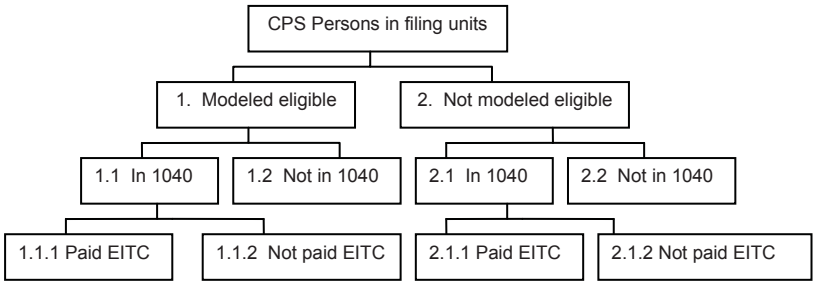
Four general results are possible when attempting to match an EITC-eligible CPS ASEC respondent to the IRS tax return data:

- A match occurs between the two databases, and the respondent received EITC from the IRS (Figure 3–Box 1.1.1)
- A match occurs between the two databases, and the respondent did not receive EITC from the IRS (Figure 3–Box 1.1.2).

- No match occurs, indicating the respondent did not file a return (Figure 3–Box 1.2).
- No match occurs because the respondent did not obtain a PIK. This group was excluded prior to matching the datasets.

Figure 3 displays a tree detailing all of the possible combinations of CPS ASEC eligibility status, filing status, and claimant status.

Figure 3. Potential Results of Matched Datasets



At this point, a participation rate could be calculated by dividing the number of records in box 1.1.1 by the total number of records from box 1. Computing participation rates at this step assumes CPS ASEC respondents: accurately report their incomes to the Census Bureau, accurately report their incomes to the IRS, and follow modeled filing behavior.[27]

The taxpayer participation rate that results is 63 percent, which is substantially lower than previous estimates by other researchers. Our assumptions do not always hold and warranted further investigation.

The number of returns per household based on IRS filing units and on modeled survey filing units was tabulated. It was clear that the modeling predicted too many filing units in some households, and too few filing units in other households. Even in households where one unit was modeled and only one claimed, the modeling may have assigned the tax head to a person other than the claimant. The difficulty of modeling behavior was clear: the modeled filers did not always match the actual filers, and the assignment of qualifying children was often puzzling and sometimes illogical. The

[27] Assumptions 1 and 2 also imply that filers report their incomes to both agencies using similar income concepts. However, this may not be the case. For example, a respondent could report gross wages to the Census but taxable wages to the IRS.

mismatch of expected filing units (thus income and credit amounts) and actual filing units demanded revision of the modeled units.

Without adjustments to correct for instances when any of the three assumptions do not hold, the estimated participation rate will be in error (and understated). Therefore, the initial set of EITC-eligible respondents was reevaluated to ensure that income reported to both agencies was within reasonable agreement. In addition, households that had at least one unanticipated tax unit were investigated to ensure that legal taxpayer behavior was incorporated into the participation estimate.

Realigning EITC Eligibility

The data were evaluated to determine how well the model predicted actual filing behavior. The modeled cases were realigned per the actual return where deemed appropriate. Some of these modifications were made to best utilize the IRS data. Two objective adjustments were made to address unpaid claims and married separate filers. The special EITC extract transmitted to the Census Bureau included variables on the amount of EITC claimed and the amount actually paid. The flag used to this point had included the presence of any EITC information on the return. At this point, only cases with EITC paid were used in the analysis. This seemed appropriate to do when computing the participation of persons who received the credit. The second adjustment removed eligibility for persons who filed married separate returns. EITC rules prohibit married separate filers from claiming the credit. Using the household information in the CPS ASEC, modelers are unable to predict which household will choose to file married jointly or married separate. The filing category variable transmitted by the IRS was used to make this adjustment. These two adjustments reduced the number of eligible modeled filers by 225,000, from 18.27 million to 18.05 million.

Additional, more subjective, changes were made to modeled filers. These adjustments were made following IRS 1040 and EITC instructions, with consultation from W&I staff. Clerical review of the households with one or more EITC claimants per model or per IRS revealed that many households claimed children differently than the model had predicted, or did not claim children present in the survey at all. Additionally, childless households in the survey filed returns with EITC qualifying children. More analysis is needed to determine whether the unclaimed children in some returns and the unanticipated children in other returns balance out when viewing the national survey results.

The cases that could be altered were those where children were modeled to one adult but claimed by another. A similar situation occurred when one adult was modeled and another in the household was the claimant. The filing status and qualifying child assignment for these units were flagged and adjusted manually. These adjustments resulted in a 231,000 increase in eligible returns, bringing the total to 18.28 million. The adjustments also affected the numerator of the participation rate, with a 337,000 increase, bringing the total of matched eligible units who received the credit to 11.96 million.

The following scenarios provide specific examples of when eligibility was reassigned.

- Two unmarried adults live in the same household with one child. Adult A is the known parent of the child and was modeled as eligible for EITC. Adult B actually claimed EITC with one qualifying child. In the 2006 CPS ASEC, information identifying the second parent was not collected; however, if the second adult was a known relative (grandparent, aunt, etc.) of the child, then eligibility was reassigned to Adult B (if income thresholds permitted). If Adult A claimed childless EITC, he or she was not counted as an eligible claim due to the rule that disallows taxpayers from claiming EITC if they have a qualifying child who was claimed for EITC on another person's tax return.
- Same situation as 1, except Adult A is the known parent of two children in the household. If both adults claimed EITC using one qualifying child, both claims were counted as legal claims.
- Same situation as 2, except Adult B is not related to Adult A but was within 20 years of Adult A. It was assumed that Adult B was cohabitating with Adult A and was the other parent of the child and therefore eligible to claim the credit. EITC eligibility was reassigned to Adult B.

Changes to this point were made based on a person mismatch that shifted the filing assignment within the unit. There were still many cases where qualifying child(ren) were modeled differently than the administrative record indicated. The data were reviewed to determine the best approach to address these seemingly eligible units. In the majority of these cases, children were assigned to a filer who did not actually claim them. If the IRS child at home exemption flag verified that no children (or fewer than

modeled) were present on a return, then, based on a cursory AGI test, the unit was removed from eligibility. This stage also removed eligibility in cases with overstated survey income. The incidence and extent of survey tax return income reporting discrepancies will be addressed in the future. At this time, it seemed appropriate to remove eligibility on a combination of tax return qualifying child (QC) and income data. The following rules were applied to automatically remove eligibility for households meeting these conditions:

- 0 qualifying children were modeled, no child exemptions were present, and AGI was greater than $11,750.
- 1 qualifying children was modeled, no child exemptions were present, and AGI was greater than $34,000.
- 2 qualifying children were modeled, 2+ child exemptions were present, and AGI was greater than $38,000.
- 2 qualifying children were modeled, 1 child exemption was present, and AGI was greater than $32,000.

These changes impacted both the number of eligible and number of paid returns for our analysis. The number of eligible cases fell from 18.28 million to 14.99 million, a 3.29-million reduction. The number of paid returns fell from 11.96 million to 11.29 million, a 668,000 reduction. This reduction of 668,000 taxpayers seems to indicate that the Census Bureau algorithm may not have been precise enough when identifying cases that did not appear eligible. This algorithm will be investigated for future improvement.

Future research projects will investigate households that did not have a tax unit identified as EITC-eligible, but were paid EITC, as there are likely to be cases where EITCeligibility was incorrectly modeled. We have preliminarily identified tax units who reported no earned income to the survey but the tax return reported wage income. If all other information between the two agencies agrees and income amounts are within EITC tolerances, these tax units could have their eligibility reassigned. To facilitate this effort, an approximation for gross wages from W–2s will be computed and transmitted to the Census Bureau for future work. The W–2 will confirm the existence of earned income, and gross wages from the W–2 will be used to determine if income discrepancies between the two agencies are a result of CPS ASEC respondents reporting gross incomes to the Census Bureau and taxable incomes to the IRS.

Results

Participation Estimates after Adjustments

After the filing units were realigned, the taxpayer participation rate (TPR) was computed by dividing the number of modeled units who received the EITC per IRS by the total number of modeled units. This analysis used the CPS ASEC-modeled eligibles as the base and used the record check methodology to assess participation. Note that W&I Research originally predicted eligibility for 13,393 records, but fewer than 9,000 were used in the TPR computations. 15 percent (2,036 records) were omitted because no PIK was assigned to the survey record. 9 percent (1,022 records) of those remaining were omitted because their survey data were fully imputed. Of those left, 18.1 percent (1,873 records) were omitted because their earnings amounts were imputed.

Of the 14,081 EITC recipients PIKed by the Census Bureau, only 76 percent were used in the TPR computation.[28] These records will be evaluated in the future to determine why no eligibility was modeled given the CPS ASEC information. It is anticipated that the presence and assignment of EITC-qualifying children and income reporting differences between the agencies will be important factors.

W&I Research estimated 19.05 million tax units (single persons or families) eligible to claim $31.4 billion for TY 2005. The total counts of eligible taxpayers in the following tables sum to just under 15 million (14,988,890), a reduction of about 4 million from those originally modeled (because of the adjustments previously described). The true number of taxpayers eligible to receive EITC for TY 2005 has not been determined as of the writing of this report (due to the issues previously discussed that still require investigation), but it is likely to be larger than the 14.9 million reported in the following tables and less than the originally modeled 19.05 million. The reader should realize that future adjustments to the population identified as eligible for EITC will alter the participation estimate. Additionally, the number of nonclaimants will likely increase with future revisions (but that does not necessarily mean the participation rate will decrease if the number of claimants increases proportionately).

Based on the results of the match and subsequent adjustments, an estimated 11.3 million of the remaining 14.9 million taxpayers included in the

[28] Based on the TY 2001 NRP audit results of individual tax returns, about 65 percent of TY 2001 EITC claimants were EITC-eligible. In TY 2002, tax law changes went into effect that may have reduced the percentage of taxpayers who were noncompliant.

analysis were paid EITC, resulting in a participation rate of 75 percent (+/–2 percent). (Again, the reader should note this count of 11.3 million likely understates the true number of eligible recipients and will be revised.) Table 8 provides a breakout of the number of eligible tax units who filed a tax return. Most of the tax units who were identified as eligible for EITC filed a tax return (84 percent) and were paid EITC (75 percent). Taxpayers who do not file a tax return account for about two-thirds of nonparticipants (2.4 million out of 3.7 million nonclaimants).

Table 8: Census Eligible Population, by Filing Status

Filing Status	EITC Status	Eligible Count	Eligible Percent
Filed	Paid	11,289,390	75.3%
	Not Paid	1,300,100	8.7%
Did Not File	Not Paid	2,399,400	16.0%
Total		14,988,890	100.0%

Source: TY2005 IRS–CPS ASEC Exact Match

Table 9 provides the participation rate by the number of qualifying children present in the tax unit. Not surprisingly, the participation rate increases as the number of qualifying children increases (which is directly tied to the number of dollars a taxpayer is eligible to receive). While taxpayers with no qualifying children had the lowest participation rate, it should be noted that there are more taxpayers with qualifying children who are not participating (2.3 million) than taxpayers without qualifying children (1.4 million). Taxpayers with qualifying children represent about 60 percent of the eligible nonclaimants. The participation rate for taxpayers with qualifying children is estimated to be 81 percent, which compares well to Scholz's estimate of 80 percent to 86 percent for TY 1990 (when there was no credit for childless workers).

Table 9: Participation Rate, by Number of Qualifying Children

Qualifying Children	Observations	Number Paid EITC (Weighted)	Number Modeled Eligible (Weighted)	Eligible Nonfilers (Weighted)	Participation Rate with Margin of Error
0	1,256	1,738,125	3,124,484	1,029,500	55.6% +/−3%
1	2,628	3,803,345	5,171,023	658,870	73.6% +/−2%
2+	3,526	5,747,930	6,693,383	713,490	85.9% +/−2%
Total	7,410	11,289,400	14,988,890	2,401,860	75.3% +/−2%

Source: TY2005 IRS–CPS ASEC Exact Match

The IRS sends eligibility notices to taxpayers who file tax returns, appear eligible for EITC, but do not claim the credit. In TY 2005, there were 622,000 of these letters sent to taxpayers notifying them of potential eligibility. In an effort to reduce mailing costs in Calendar Year 2006, approximately 100,000 to 140,000 taxpayers who normally would have been notified of their eligibility were not mailed a notice.[29] There likely would have been about 720,000 to 760,000 notices mailed in TY 2005 if the eligibility notices not been suppressed. Additionally, some taxpayers who appear eligible for EITC, but do not claim EITC, are not sent a notice due to the following issues:

- Current or past compliance issues related to EITC or dependents
- Uncertainty in accurately determining EITC eligibility based on return information[30]
- Certain types of income are present
- Taxpayer specifically indicates on the tax return that he or she does not want to receive EITC (i.e., for religious reasons).

Table 10 shows the number of taxpayers remaining eligible for EITC as the EITC rules are applied to the tax return (applying the rules in a different order would result in different intermediate counts). Steps 1 to 11 remove taxpayers who do not pass the eligibility rules, and steps 13–18 identify reasons why an eligibility notice was not sent to the taxpayer. W&I Research estimates an additional 700,000 taxpayers in TY 2005 who could have been sent a notice but did not receive one (step 12). When the number of notices that do not result in a claim is added to the number of additional notices that could be sent, the total is around one million returns (step 11). This number corresponds to the estimated number of filer, nonclaimants (1.3 million) derived from the Exact Match. It seems that the number of eligible nonclaimant filers identified as a result of the match is too high, and could indicate that the participation rate is understated. For future studies, the IRS will transmit a list of taxpayers who received one of the eligibility notices and the list of the taxpayers who could have received the notice. If the taxpayer is not in the list of payees or the list of the CP–09/27 notices (actual and suppressed), he or she will be deemed not eligible for EITC, and, if the taxpayer is in the list of CPS ASEC tax units eligible for EITC, it is anticipated that he or she will be removed from that list.

[29] Taxpayers who computer-prepared and printed their returns were not sent the eligibility letter.

[30] For example, taxpapers with dependent children over age 18 are not sent a notice because the IRS does not know if the child was enrolled in school. Dependents over 18 and not enrolled in school are not qualifying children for EITC (unless the dependent is totally and permanently disabled).

Table 10: Step by Step Results of EITC Eligibility Determination of Nonclaimants

Step	Tax Returns Eliminated	Tax Returns Remaining	Filter Applied
Start	—	133,646,046	—
1	126,741,070	6,904,976	Taxpayers with invalid primary TIN; taxpayers who used an ITIN; Earned Income or AGI exceeded $37,262; taxpayers with investment income that exceeded thresholds (preliminary definition of investment income); taxpayers who claimed EITC; taxpayers using a filing status of Married Filing Separate; taxpayers with no earned income; taxpayers claimed as a dependent on another taxpayer's return
2	4,555,277	2,349,699	Taxpayers under 25 or over 64 with no dependents
3	338,013	2,011,686	Taxpayers with Earned Income exceeds QC thresholds (preliminary definition of QC)
4	348,913	1,662,773	Taxpayers with dependents over age 23 and do not meet age and income requirements
5	110,480	1,552,293	Taxpayers with investment income (revised definition) exceeding thresholds
6	12,616	1,539,677	Taxpayers with Form 2555
7	74,462	1,465,215	Taxpayers residing in U.S. territories
8	168	1,465,047	Taxpayers who had EITC manually adjusted during processing
9	8,372	1,456,675	Taxpayers whot indicate they are not eligible for EITC and are not subject to self-employment tax
10	33,970	1,422,705	Taxpayers using Form 1040NR
11	366,441	1,056,264	Taxpayers paid EITC after return processing
12	267,696	788,568	Taxpayers who received an eligibility notice
13	375,462	413,106	Taxpayers with a Return Processing Code of B present
14	204,778	208,328	Taxpayers who completed their returns electronically, but printed and mailed the returns
15	113,468	94,860	Taxpayers with all qualifying children ages 19 to 23
16	11,181	83,679	Taxpayers who filed their returns in 2007 (late filers)
17	272	83,407	Taxpayers who filed as Married Filing Joint but did not report a secondary SSN/TIN
18	35	83,372	Taxpayers with a criminal investigation indicator

Source: W&I Research CP–09/27 Eligibility Algorithm applied to TY 2005 IRTF

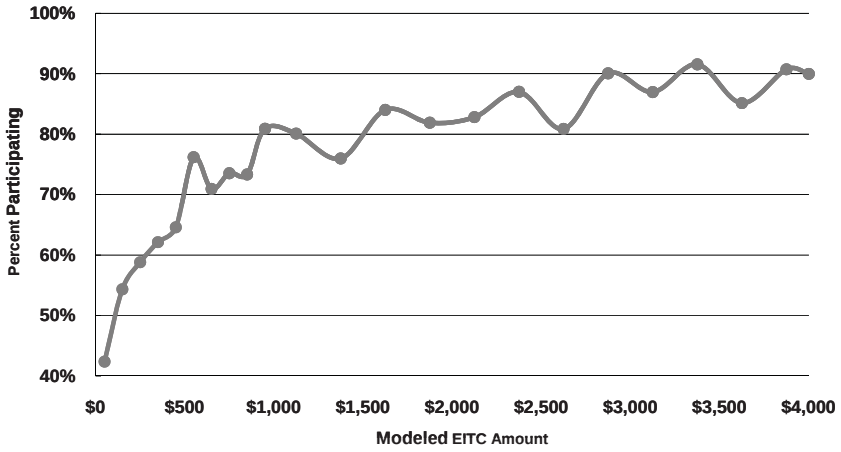
Table 11 provides the participation rate by grouped values of the EITC amount that taxpayers were estimated to be eligible to receive. The table clearly demonstrates that the participation rate increases with the amount of EITC. Taxpayers eligible for less than $100 were paid EITC less than 50 percent of the time, while taxpayers eligible for amounts greater than $4,000 were paid 90 percent of the time. The only dollar range that appeared to have a lower level of participation than the previous dollar range was the $600–$699 group, but that difference was not statistically significant. Given the relative small sample size of the $500–599 group, the $500–$599 group may have an overstated participation estimate due to sampling variability. Figure 4 shows the participation rate versus the modeled EITC amount. Increases of $100 in the EITC amount appear to have a larger impact on participation when the value is less than $600.

Table 11: Participation Rate, by EITC Amount

Modeled EITC Amount	Number Paid EITC (Weighted)	Number Modeled Eligible (Weighted)	Participation Rate
$1–$99	398,515	940,524	42%
$100–$199	606,724	1,116,412	54%
$200–$299	506,430	861,120	59%
$300–$399	782,476	1,259,337	62%
$400–$499	261,608	405,019	65%
$500–$599	95,075	124,788	76%
$600–$699	208,555	294,021	71%
$700–$799	77,790	105,787	73%
$800–$899	188,804	257,437	73%
$900–$999	273,046	337,551	81%
$1,000–$1,999	2,257,904	2,792,087	81%
$2,000–$2,999	2,743,983	3,287,461	84%
$3,000–$3,999	1,297,015	1,473,187	88%
$4,000+	1,279,818	1,422,511	90%
Not Originally Modeled Eligible	311,650	311,650	—
Total	11,289,390	14,988,890	75%

Source: TY2005 IRS–CPS ASEC Exact Match

Figure 4. Participation Rates by Modeled EITC Credit



SOURCE: TY 2005 IRS-CPS ASEC Exact Match.

Table 12 provides a breakout of taxpayer participation by marital status/gender and number of qualifying children. Three groups exceeded the national participation rate of 75 percent—single females with one or two (or more) qualifying children (80 percent and 90 percent, respectively) and married filers with two (or more) qualifying children (84 percent). Single, male taxpayers with no qualifying children had the lowest participation rate (48 percent). Single, male taxpayers lagged single, female taxpayers in each qualifying child category by margins of 13 percent to 21 percent.

Table 12: Participation Rate, by Marital Status and Qualifying Children

Marital Status	Qualifying Children	Observations	Number Paid EITC (Weighted)	Number Modeled Eligible (Weighted)	Participation Rate with Margin of Error
Married	0	156	207,571	342,554	61% +/−10%
	1	778	955,434	1,502,750	64% +/−5%
	2+	1,615	2,464,920	2,939,910	84% +/−3%
Male	0	545	683,728	1,435,040	48% +/−5%
	1	356	497,891	740,613	67% +/−7%
	2+	220	287,460	413,963	69% +/−9%
Female	0	555	846,826	1,346,890	63% +/−5%
	1	1,494	2,350,020	2,927,660	80% +/−3%
	2+	1,691	2,995,550	3,339,510	90% +/−2%
Total		7,410	11,289,390	14,988,890	75% +/−2%

Source: TY2005 IRS–CPS ASEC Exact Match

Table 13 provides the participation rate by age of the taxpayer and number of qualifying children. Taxpayers under age 25 are eligible only if they have qualifying children; therefore, it is not surprising that the participation rate for those under 25 (81 percent) is in the participation range for taxpayers with qualifying children (79 percent to 83 percent). For taxpayers older than 44, the percentage of eligible taxpayers with no qualifying children increases for each age group, which likely is contributing to decreasing participation rates (eligible taxpayers with no qualifying children have a lower participation rate than taxpayers with qualifying children).

Table 13: Participation Rate, by Age of Taxpayer

	Age Category	Observations	Number Paid EITC (Weighted)	Number Modeled Eligible (Weighted)	Participation Rate with Margin of Error
0 QC	<25	4	3,701	6,682	56% +/−72%
	25-34	469	724,386	1,271,700	57% +/−5%
	35-44	226	293,062	558,855	52% +/−8%
	45-54	304	413,169	728,210	57% +/−7%
	55+	253	303,808	559,033	54% +/−8%
1 QC	<25	431	698,733	881,840	79% +/−5%
	25-34	743	1,155,180	1,497,310	77% +/−4%
	35-44	694	952,075	1,317,210	72% +/−5%
	45-54	564	764,292	1,096,190	70% +/−5%
	55+	196	233,066	378,479	62% +/−10%
2+ QC	<25	310	556,954	670,469	83% +/−4%
	25-34	1,349	2,285,480	2,639,410	87% +/−4%
	35-44	1,288	2,021,810	2,310,750	88% +/−4%
	45-54	474	734,497	881,871	83% +/−4%
	55+	105	149,186	190,889	78% +/−4%
All Groups	<25	745	1,259,390	1,558,990	81% +/−6%
	25-34	2,561	4,165,040	5,408,420	77% +/−3%
	35-44	2,208	3,266,950	4,186,810	78% +/−3%
	45-54	1,342	1,911,960	2,706,270	71% +/−5%
	55+	554	686,061	1,128,400	61% +/−11%
	Total	7,410	11,289,400	14,988,890	75% +/−2%

Source: TY2005 IRS-CPS ASEC Exact Match

The Data Integration Division of the U.S. Census Bureau created unique geographic divisions of the United States to determine if there was a difference in EITC participation rates in different parts of the United States (Table 14). The definitions of the geographies follow:

- East Central=Ohio, Indiana, Illinois, Missouri, West Virginia, Kentucky, Tennessee, Arkansas
- East Coast=New York, New Jersey, Pennsylvania, Delaware, Maryland, District of Columbia, Virginia, North Carolina
- New England=Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut
- North Central=Michigan, Wisconsin, Minnesota, Iowa, North Dakota, South Dakota, Montana, Idaho
- Southeast=South Carolina, Georgia, Florida, Alabama, Mississippi, Louisiana
- Southwest=Nebraska, Kansas, Oklahoma, Texas, Colorado, New Mexico, Arizona, Utah
- West=California, Washington, Oregon, Idaho, Nevada, Alaska, Hawaii

The States were grouped in this fashion in order to create a division that contained the States from the Southeast region of the U.S. Previously defined Census Bureau divisions divide the States from the Southeast into three different divisions. The Southeast contains a large proportion of the EITC claimants, and the theory was that the Southeast might exceed other areas in participation.[31] The West lags the national participation rates in each of the three qualifying children groups, with the largest lag in the zero qualifying children segment (46 percent versus 56 percent nationally). Previous work by W&I Research has found the West to lag in participation. The same work found the South region (as defined by the Census Bureau) to have increased levels of participation, which was not found in this effort.[32]

[31] The residents of these six States submit 19 percent of all EITC claims (and receive 21 percent of all EITC dollars), but the total number of tax returns from these States represents only 14 percent of all tax returns filed. About 17 percent of all tax returns report EITC. In these States, 23 percent of tax returns report an EITC claim. Mississippi has the highest percentage (32 percent) of tax returns reporting EITC.

[32] The South Region includes the South Atlantic, East South Central, and West South Central divisions.

Table 14: Participation Rate, by Geographic Area

Geographic Area	Observations	Number Paid EITC (Weighted)	Number Modeled Eligible (Weighted)	Participation Rate with Margin of Error
East Central	1,134	2,039,670	2,681,200	76% +/−3%
East Coast	1,148	2,070,180	2,719,100	76% +/−3%
New England	646	437,751	577,134	76% +/−7%
North Central	996	908,616	1,201,720	76% +/−4%
Southeast	875	1,792,560	2,391,510	75% +/−3%
Southwest	1,334	2,236,790	2,887,850	78% +/−3%
West	1,277	1,803,830	2,530,370	71% +/−3%
Total	7,410	11,289,400	14,988,890	75% +/−2%

Source: TY2005 IRS–CPS ASEC Exact Match

AGI and earned income are used to determine the amount of EITC each taxpayer is entitled to receive. The EITC amount is based on earned income when AGI is less than the income amount associated with the phaseout range of the program, which varies by number of qualifying children. The EITC amount is based on AGI and earned income when the AGI amount is in the phaseout range. A plot of earned income on the x-axis and the EITC benefit on the y-axis provides a graph that looks like a pyramid with a plateau at the top. As earned income increases from zero, the amount of the credit also increases, and that income range is termed the phase-in range (left side of the pyramid). At a certain income, the amount of EITC is constant (the maximum benefit range), even with increases in income (plateau area of the pyramid). Finally, further increases in income result in a decreased amount of EITC (phaseout range) until the income exceeds the maximum amount of income to qualify for EITC (right side of the pyramid). Table 15 provides the income ranges for each of the three income ranges, by the number of qualifying children.

Table 15: Incomes Ranges for Phase-In, Maximum Benefit, and Phaseout Groups

	Phase-In Income Range	Max Benefit Income Range*	Phase-Out Income Range*
0 QC	$1–$5,199	$5,200–$6,549	$6,550–$11,749
1 QC	$1–$7,799	$7,800–$14,399	$14,400–$31,029
2+ QC	$1–$11,049	$11,050–14,399	$14,400–$35,262

*For taxpayers who file as Married Filing Joint, the income ranges are increased by an additional $2,000 in the max benefit and phaseout ranges.

Source: TY 2005 IRS Publication 596, Earned Income Credit

Table 16 provides participation estimates for taxpayers in the three income ranges (phase-in, maximum benefit, phaseout) related to EITC. Taxpayers in the phase-in income range show lower levels of participation when compared to taxpayers in the maximum benefit and phaseout income ranges. This finding is true for all qualifying children categories. There does not appear to be any significant difference in participation between the maximum benefit and phaseout income groups. Taxpayers in the phase-in group may not have a filing requirement, and given their low incomes, are likely to have had less income tax withheld and have less incentive to file a return to receive a refund of withholding—this is especially true of taxpayers with no qualifying children. This may explain why taxpayers with incomes in the phase-in range lag the other income groups in participation.

Table 16: Participation Rate, by Benefits Phase

	Benefits Phase	Observations	Number Paid EITC (Weighted)	Number Modeled Eligible (Weighted)	Participation Rate with Margin of Error
0 QC	Phase-In	573	662,280	1,426,270	46% +/-5%
	Max Benefit	135	218,605	342,863	64% +/-10%
	Phaseout	548	857,240	1,355,340	63% +/-5%
1 QC	Phase-In	528	706,802	1,047,710	68% +/-6%
	Max Benefit	472	745,863	947,226	79% +/-5%
	Phaseout	1,628	2,350,680	3,176,090	74% +/-3%
2+ QC	Phase-In	764	1,211,970	1,513,720	80% +/-4%
	Max Benefit	378	655,275	735,325	89% +/-4%
	Phaseout	2,384	3,880,670	4,444,340	87% +/-2%
All Groups	Phase-In	1,865	2,581,060	3,987,710	65% +/-3%
	Max Benefit	985	1,619,740	2,025,410	80% +/-3%
	Phaseout	4,560	7,088,590	8,975,770	79% +/-2%
	Total	7,410	11,289,400	14,988,890	75% +/-2%

Source: TY2005 IRS–CPS ASEC Exact Match

Table 17 provides a breakout of modeled AGI for tax units modeled as eligible who did not file a tax return. The vertical boxes contain the tax units who would not have a filing requirement based solely on their modeled AGIs.[33] The counts in the boxes are summed in the row "AGI below filing

[33] Filing requirements are based on gross income and not AGI. Gross income includes gross (rather than net) business income. The use of modeled AGI is a close approximation to gross income as it only contains one subtraction from income (one-half of self-employment tax). However, the modeled AGI includes net income, not gross income. About 11 percent of the eligible nonfilers had self-employment income. Income thresholds were obtained from the TY 2005 Form 1040 Instructions, page 12. The income threshold associated with taxpayers under age 65 was the income threshold used for the three filing statuses.

requirement" to give the total, estimated number of tax units by modeled filing status without a filing requirement. Roughly 60 percent of the nonfilers did not have a filing requirement, with about 1.2 million having less than $8,000 in AGI. If the respondent had an AGI of less than $8,000, it is understandable that the respondent did not file a tax return, given the respondent's expected benefit of filing a return (small refund of withholding, if any) and costs associated with filing (paid preparer).

Table 17: Modeled AGI Categories for Eligible Non-filers (Weighted Modeled Returns) [4]

AGI Amount	Single	Head of Household	Married Filing Joint	Total
Negative	236	0	0	236
$0	21,941	7,488	5,495	34,923
$1–$1,000	151,810	37,053	100,170	289,033
$1,001–$2,000	90,901	16,556	62,117	169,574
$2,001–$3,000	66,161	19,722	47,530	133,413
$3,001–$4,000	56,346	10,598	33,737	100,681
$4,001–$5,000	88,445	2,171	51,157	141,773
$5,001–$8,000	231,081	61,048	99,467	391,596
$9,001–$10,000	115,793	40,021	57,349	213,163
$11,001–$16,000	116,694	89,926	146,431	353,051
$16,001+	0	304,580	267,340	571,920
AGI below filing requirement	684,744	187,169	597,958	1,469,871
Total	939,408	589,162	870,793	2,399,363

Source: TY 2005 IRS–CPS ASEC Exact Match

[4] Nonfilers with modeled AGI equal to or less than zero were not included in the group with no filing requirement because they reported negative income, which may be obscuring a filing requirement.

Table 18 reports the participation rates by major industry.[34] Taxpayers in the education/health services, financial activities, and wholesale/retail trade had higher levels of participation, while taxpayers in the construction and information industries had lower levels of participation. Unmarried females were the predominant group in each of the three highest participating industries (education/health at 77 percent, financial services at 65 percent, and wholesale trade at 57 percent). Unmarried females make up 51 percent of the 14.9 million taxpayers eligible for EITC.

[34] See Appendix A of the March 2006, ASEC technical documentation for industry coding details <http://www.census.gov/apsd/techdoc/cps/cpsmar06.pdf>.

Table 18: Participation Rate, by Industry

Industry	Number Paid EITC (Weighted)	Number Modeled Eligible (Weighted)	Participation Rate
Agriculture, forestry, fishing, hunting	149,017	188,299	79%
Construction	574,630	915,435	63%
Manufacturing	933,604	1,169,684	80%
Wholesale and retail trade	1,557,299	1,917,801	81%
Transportation and utilities	368,115	461,236	80%
Information	102,778	162,026	63%
Financial activities	511,599	628,566	81%
Professional and business services	825,333	1,124,690	73%
Educational and health services	2,476,930	2,952,245	84%
Leisure and hospitality	1,176,963	1,524,511	77%
Other services	581,188	823,724	71%
Public administration	296,196	372,426	80%
Other	1,735,741	2,748,247	63%
Total	11,289,390	14,988,890	75%

Source: TY2005 IRS-CPS ASEC Exact Match

Conclusions

Employing an Exact Match methodology that relies solely on information reported to the Census Bureau and that does not incorporate information from tax return filings underestimates the participation estimate because of income underreporting. Taxpayer filing behavior that cannot be anticipated prior to a comparison of tax return filings also causes the participation rate to be underestimated. The taxpayer participation rate that resulted after appropriate adjustments were made to the set of eligibles is in alignment with previous participation estimates produced by Scholz, GAO, and W&I Research. The characteristics of modeled eligible nonclaimants seem to confirm the validity of the methodology employed in this project as nonclaimants were generally nonfilers. And tax units modeled as EITC-eligible were more likely to claim the credit as the amount of the modeled credit increased.

The taxpayer participation rate appears to be relatively stable over time (the current estimate of participation for taxpayers with children is within the range Scholz estimated for TY 1990, and the national estimate is the same as GAO's estimate for TY 1999). Sustained substantial increases in the participation rate may be difficult to achieve if filing requirements remain

the same, and/or credit amounts are not increased to induce consistent filing by eligible nonfilers.[35]

Improving participation among taxpayers with smaller credit amounts and/or no filing requirement will be difficult. Taxpayers with no filing requirement who are unaware of their EITC eligibility may determine that the costs of filing a tax return outweigh the benefits of filing a tax return. Migrating low-income taxpayers to low-cost, return preparation options may improve participation rates.

Taxpayers with no qualifying children had lower rates of participation than taxpayers with qualifying children, but the majority of the nonclaimant population appeared to have at least one qualifying child. Viewing the nonclaimant population as primarily childless workers does not appear to be an accurate assessment of the population.

The participation rate estimate will likely change after the inclusion of several planned improvements to the methodology, scheduled to occur late in 2009 (described in the following section). It is anticipated that these improvements will produce a higher participation estimate. The planned improvements will also allow for a more accurate estimate of the total number of eligible taxpayers and eligible participants/nonparticipants.

Limitations and Future Improvements

Correctly Identifying Population Eligible To Receive EITC

The subsample identified as EITC-eligible from CPS ASEC excludes some individuals/families who are eligible for EITC and includes tax units who likely are not EITC-eligible. For an Exact Match methodology to be successful, the modeling must not misclassify a tax unit as ineligible when the tax unit actually is eligible to receive EITC. The misclassified tax units would generally fall into boxes 1.1.2 and 2.1.1 shown in Figure 3. Box 1.1.2 has been investigated, leaving only box 2.1.1 for future study.

[35] TY 2007 saw a 1.6-million increase in the number of returns claiming EITC. Most of the increase in claims was due to the general increase in the total number of returns that resulted from the Economic Stimulus Program in TY 2007 that sent rebate checks to qualifying taxpayers. Taxpayers were induced to file for the stimulus payments and also qualified for EITC. (There was an increase of about 300,000 first-time EITC claimant taxpayers over earlier tax years.) It is too soon to know whether claims in TY 2008 drop back to TY 2006 levels.

Income Discrepancies

One of the key requirements for EITC eligibility is meeting the income requirements (having earned income greater than zero and having earned income and AGI both less than certain thresholds). When reporting income to the Census Bureau, it is possible that a respondent will provide rounded income to the Census Bureau (unless the respondent has tax return/income reporting documents in hand). In most situations, the rounding of income by a survey respondent is not a problem when attempting to determine eligibility (it does cause issues when trying to estimate the dollar amount). However, in cases where the taxpayer is very close to the EITC thresholds, rounding can cause significant issues. For example, suppose a married couple earns $32,000 and has one qualifying child. When asked by the Census Bureau official how much they earned, they round up and report $35,000 (an overstatement of less than 10 percent). The cutoff for a married couple with one qualifying child was $33,030 in TY 2005, meaning that the couple was eligible for EITC but would not be classified as eligible because $35,000 exceeds $33,030.

CPS ASEC respondents reporting gross income to the Census Bureau and taxable income to the IRS have a similar impact on eligibility determination to rounding. Many workers contribute to retirement plans (401K/403B/TSP/SEP), enroll in employer-offered health insurance plans, and enroll in medical/dependent care flexible spending accounts. These payroll deductions are deducted from gross wages before withholding is applied, as taxable wages are equal to gross wages minus these deductions. If someone had $35,000 in gross wages and accurately reported that income to the Census Bureau, but taxable wages were $30,000 (as a result of health insurance deductions and tax-deferred investments), and he or she had one qualifying child, he or she would also be wrongly categorized as not eligible for EITC through appearing to have too much income.

Another source of misreporting arises when income is reported by another person in the household. When a Census Bureau employee obtains information about a household, the interviewer obtains the information from the reference person. In households with multiple families, the householder may not have complete knowledge of the earnings of other household members or may not feel comfortable sharing information about others in the

household. If the householder does not report income for his or her relatives or housemates, or is unable to provide accurate estimates of that income, then accurately determining EITC eligibility for these persons will be difficult and prone to error.

An obvious source of income mismatches results from reporting noncompliance. In these instances, the taxpayer underreports income to the IRS and/or overstates adjustments to income and self-employment expenses to generate a smaller tax liability. This misreporting could result in the taxpayer appearing eligible for EITC to the IRS, but the taxpayer would not have been modeled as eligible (assuming the taxpayer accurately reported income to the Census Bureau). In instances where the difference in income is large (for example, $40,000 versus $20,000), it may be easier for an outside party to assume the difference is a result of noncompliance (especially if the income source is self-employment), but, in instances where the difference is relatively small (for example, $35,000 versus $30,000), it is more difficult to make the assumption that the difference is noncompliance because of the possibilities previously discussed.

In future studies, the IRS will transmit an estimate for gross wages, in addition to taxable wages, to study the impact the different income concepts have on the eligibility determination. Analysts will correct for instances where the difference in income is a result of reporting gross wages to the Census Bureau. The Census Bureau will also explore the impact of self-reported versus proxy-reported income on accurately determining EITC eligibility (especially as it relates to taxpayers who show no earned income in the survey data but report wage income on their tax returns) and will provide recommendations on corrective action.

Inability To PIK all CPS ASEC Respondents

This research assumes that all of the CPS ASEC respondents who do not have a PIK assigned are missing at random. This may be a valid assumption for respondents who are legal U.S. residents, but it clearly is not valid for U.S. residents who are not residing in the U.S. legally.

The IRS provides ITINs to noncitizens without SSNs (residing in the U.S. and abroad) to properly track their tax accounts. In future studies, the IRS will provide the Census Bureau with a population file of ITIN applicants residing in the U.S. so that the Census Bureau will be able to identify U.S. residents modeled as EITC-eligible who have an ITIN and remove them from the set of tax units modeled eligible for EITC. Variables included in the transmission would include name, address, country of birth, date of birth,

and gender. With these variables, the Census Bureau will assign unique link identifiers and use these cases in the analysis. The removal of this population from the set identified as eligible will not correct the whole issue of modeling U.S. residents without SSNs as eligible for EITC (because not all U.S. residents without an SSN apply for an ITIN), but it is a step in the right direction. The number of persons with an ITIN is estimated to be about 15 million (not all ITIN users reside in the U.S.).

Table 19 provides a summary of the limitations encountered by the Exact Match method and the projected impact of each limitation on the estimated participation rate. No actual percentages are reported in the table because the issues require research in future work. Once impacts on the participation rate are quantified, the participation rate will be modified accordingly. Most of these limitations are likely to cause the participation estimate to be understated.

Table 19: Limitations to the Exact Match Methodology and Impact on Participation

Limitation	Projected Impact on Participation Estimate
Income Discrepancies —Reporting Gross Income versus Taxable Income	Negative
Income Discrepancies —Third Party Income Reporting	Negative
Income Discrepancies —Reporting Noncompliance	Neutral
Inability to PIK All CPS ASEC Respondents —U.S. Residents without an SSN	Neutral
Inability to PIK All CPS ASEC Respondents —U.S. Residents with an SSN	Unknown
Modeling Ineligible Tax Units as Eligible for EITC —Inappropriate Assignment of QC to Tax Unit	Negative

Income Discrepancies—Reporting Gross Income Versus Taxable Income—The anticipated impact of this error on the participation estimate is negative because the error is likely to be focused on taxpayers near the upper income eligibility thresholds. These taxpayers would appear to be ineligible for EITC (based on CPS ASEC gross income data) but may actually be eligible for EITC when using taxable income. Taxpayers in this income range are more likely to file a tax return (due to filing requirements) and as result of filing receive the credit. This error will cause the number of eligible taxpayers to be underestimated.

Income Discrepancies—Third Party Income Reporting—Another way the number of EITC-eligible CPS ASEC respondents is understated results when respondents are not considered eligible for EITC because they have no reported earnings in the Census Bureau data. Some of these persons have IRS tax returns showing earned income. A cursory examination of these records indicates that household responses may have been provided by the reference person rather than the filer/earner. We will investigate whether income reporting accuracy is affected by household composition, specifically in the relationship of the earner to the reference person.

Income Discrepancies—Reporting Noncompliance—Taxpayers may underreport their incomes to the IRS (to avoid a tax liability and/or to increase EITC benefits). If taxpayers also underreported their incomes on the Census Bureau survey, then our computations of EITC would be unaffected. Other income discrepancies between the agencies need to be evaluated to determine any impacts on eligibility determination and the participation estimate.

Inability To PIK All CPS ASEC Respondents—U.S. Residents without an SSN—When eligibility is computed using the CPS ASEC person records, citizenship and legal work status are unknown. This could result in modeled EITC eligibility for persons who are unable to claim the credit because they do not have SSNs. This error is addressed in the PIK process, as persons who were never assigned an SSN by the Social Security Administration do not receive PIKs. Our EITC participation analysis includes only records with PIK, so that the persons without SSNs are excluded.

Inability To PIK All CPS ASEC Respondents—U.S. Residents with an SSN—Like U.S. residents who do not have an SSN, persons who do not pass the PVS process and obtain a PIK are not included in the participation rate analysis. The reasons why a person with an SSN would not receive a PIK will be evaluated in future research. It is unclear if excluding persons with SSNs biases the taxpayer participation rate.

Modeling Ineligible Tax Units as Eligible for EITC—Inappropriate Assignment of Qualifying Children to Tax Unit—This error has been corrected by clerical review in households where multiple families reside and the child or children were assigned to the wrong family (if the families filed tax returns reporting EITC and a reassignment was appropriate). The error has not been corrected for tax units where the child was not the qualifying child of any tax unit in the household (for example, a child who resided in the household for less than the required time and may have been a qualifying child of a tax unit in another household). Nonfiling tax units may appear to be eligible for EITC but would not have a legal claim credit. These cases

would be counted as eligible nonparticipants. Other combinations of apparent eligibility and unclear legal claims will be evaluated to document and describe any resultant bias.

Appendix A: Acknowledgements

As with any project with the complexity, coordination, and magnitude that this project possessed, many people were integral to its success and completion.

They include:

- David Williams, Verlinda Paul, Patricia Lee, Lynne Morrison, Sandra Hill, Debra Holland, and Vivianne Johns from the EITC Office for their unending pursuit of obtaining a signed contract with the Census Bureau and for project management and oversight;
- Dan Beckerle, Eric Larsen, and Jeff Wilson from W&I Research for initiating the W&I effort to produce participation estimates;
- Dean Plueger for the modeling, data preparation, data transfer, and technical expertise that made the project possible;
- IRS Small Business/Self Employed Research for providing syntax to estimate EITC eligibility from CPS ASEC for TY 1996, which was updated and refined by W&I Research;
- Mark Mazur (former Director, Research, Analysis, and Statistics, IRS) and Mary-Helen Risler from IRS National Headquarters Research for providing sanity checks as the project progressed;
- Amy O'Hara from the Census Bureau for her detailed knowledge of Census procedures, her coordination of all the work done by the Census Bureau, her clerical review on weekends, and the extra effort that she provided in every phase of this project;
- Dennis Donahue and Julie Parker joined the Census Bureau in time to contribute graphics and additional analysis for this report.