UKCPR
University of Kentucky Center for Poverty Research

Discussion Paper Series
DP 2019-01

ISSN: 1936-9379

The antipoverty impact of the EITC: New estimates from survey and administrative tax records

Maggie R. Jones
Center for Economic Studies
U.S. Census Bureau

James P. Ziliak
Center for Poverty Research
and
Department of Economics
University of Kentucky

June 2020

Preferred citation
Jones, M. and Ziliak, J. (2020, June). The antipoverty impact of the EITC: New estimates from survey and administrative data. *University of Kentucky Center for Poverty Research Discussion Paper Series, DP2019-01*. Retrieved [Date] from http://ukcpr.org/research/poverty-inequality/inequality.

Author correspondence
Maggie R. Jones, margaret.r.jones@census.gov

University of Kentucky Center for Poverty Research, 550 South Limestone,
234 Gatton Building, Lexington, KY, 40506-0034
Phone: 859-257-7641. E-mail: ukcpr@uky.edu

www.ukcpr.org

The Antipoverty Impact of the EITC: New Estimates from Survey and Administrative Tax Records

Maggie R. Jones
Center for Economic Studies
U.S. Census Bureau
margaret.r.jones@census.gov

James P. Ziliak
Center for Poverty Research
and
Department of Economics
University of Kentucky
jziliak@uky.edu

June 2020

Abstract

We reassess the antipoverty effects of the EITC using unique data linking the CPS Annual Social and Economic Supplement to IRS data for the same individuals spanning years 2005-2016. We compare EITC benefits from standard simulators to administrative EITC payments and find that significantly more actual EITC payments flow to childless tax units than predicted, and to those whose family income places them above official poverty thresholds. However, actual EITC payments appear to be target efficient at the tax unit level. In 2016, about 3.1 million persons were lifted out of poverty by the EITC, substantially less than prior estimates.

* We are grateful for the assistance of Josh Mitchell on the internal Census files with TAXSIM, and to Jon Bakija for sharing his tax simulator. We benefitted from helpful discussion on early versions from Hilary Hoynes, Bruce Meyer, Jon Rothbaum, and seminar participants at the Association of Public Policy and Management Research Conference, Society of Labor Economists Annual Conference, the U.S. Census Bureau, Syracuse University, University of Illinois at Chicago, University of Kentucky, and University of Nevada-Las Vegas.

The Earned Income Tax Credit (EITC) is one of the largest and most studied antipoverty programs in the United States, with extensive research on labor supply, consumption, marriage, fertility, child achievement, and poverty alleviation.[1] Spending on the EITC exceeded $63 billion on over 25 million individuals and families in 2019[2], and the program was estimated to lift more persons out of poverty than any other safety net program for children and nonelderly working households (Hoynes and Patel 2018; Ziliak 2015a). A challenge facing nonexperimental research on the EITC is that major household surveys do not collect information on credit eligibility, receipt, or amount. Instead, EITC eligibility and amounts are simulated based on survey reports of age, family structure, earnings, income, and limited other information salient to tax liability, under the maintained assumption of 100 percent take-up conditional on eligibility. However, estimates using administrative Internal Revenue Service (IRS) data place actual take-up rates closer to 80 percent, and takeup is procyclical (Scholz 1994; Plueger 2009; Jones 2014). This suggests that estimates of antipoverty impacts from simulations assuming complete take-up may be overstated, especially in periods of economic decline. However, because the IRS pays some EITC claims that are falsely made (Marcuss et al. 2014), simulated benefits accounted for only two-thirds of aggregate EITC benefits paid (Meyer 2010), and thus survey simulations may understate the actual impact of the credit because they miss false claims. Furthermore, incorrect assignment of tax filing status and household composition may lead to over- or underestimates depending on whether families are systematically allocated close to or far away from the poverty line.

In this paper, we use a unique, internal Census dataset linking survey information from the Annual Social and Economic Supplement (ASEC) of the Current Population Survey (CPS) to

[1] See Hotz and Scholz (2003) and Nichols and Rothstein (2016) for comprehensive surveys of research on the EITC.
[2] https://www.eitc.irs.gov/partner-toolkit/basic-marketing-communication-materials/eitc-fast-facts/eitc-fast-facts

administrative IRS tax data from Form 1040, W-2 wage statements, and the EITC return file to estimate how the distribution of model-based simulated EITC benefits compares to actual administrative EITC payments for the same individuals spanning tax years 2005-2016, and the subsequent effects on after-credit poverty rates and gaps. Assessing the distributional effects of the EITC from tax simulators using survey data from the CPS ASEC versus administrative data on a leading antipoverty program is important as the CPS ASEC is the source of official income and poverty statistics in the United States (Proctor et al. 2016), as well as the Census Bureau's research supplemental poverty measure (SPM) (Renwick and Fox 2016). Moreover, the tax simulators we employ are used widely in the research and policy communities to examine effects of tax policy on labor supply, consumption, fertility, family structure, and health, and thus documenting their performance on income poverty and inequality will help gauge their likely efficacy in other behavioral domains.

Our benchmark is the IRS EITC recipient file, which records the actual EITC dollars paid to taxpayers, and provides estimates of the total dollar amounts of EITC credits circulating in the economy, regardless of whether those payments are correct. Restricting the latter to those deemed eligible provides an estimate of incorrect payments, and the ratio of those receiving IRS payment, conditional on eligibility, to those eligible who may or may not receive payment provides estimates of EITC take-up. This implies that a comparison of potential EITC payments to those eligible and those actually paid provides an estimate of the missing counterfactual of those who are eligible but do not file for the credit. Against the benchmark of actual EITC receipt, we focus on three EITC simulators: one produced by the Census Bureau as part of its annual release of the CPS ASEC (CPS model), another produced by the National Bureau of

Economic Research's TAXSIM model, and the third using Jon Bakija's tax model. We use the Bakija model on both survey-derived inputs and inputs derived from administrative data.[3]

The first exercise is to examine the full distribution of EITC payments from survey and administrative inputs across both the individual tax unit and family income distributions. The former is informative as the EITC is targeted to tax filing units and not families per se, while the latter is informative as it relates to how EITC dollars flow to the families defined by the Census Bureau for official poverty statistics. Here we find that significantly more actual EITC payments flow to childless tax units and to those whose family income places them well above official poverty thresholds than tax simulators predict. However, those seemingly "incorrect" payments still flow to low-income tax units and thus appear target efficient at the taxpayer level, and some of the discrepancy at the family level is accounted for by multi-tax-unit families.

We then examine how many persons each EITC payment model lifts out of poverty. All three survey tax simulators provide comparable estimates of the antipoverty effect of the EITC in the full CPS ASEC sample; however, compared to the benchmark sample of IRS EITC payments circulating in the economy, each model overstates the number lifted out of poverty by as much as 1.8 million persons in recent years. Even though aggregate EITC payments in the CPS, TAXSIM, and Bakija models fall short of administrative totals, they tend to allocate those payments to families closer to the poverty line. The net effect is that the estimates derived from the survey-based tax simulators predict a much greater antipoverty bite for the EITC than found in administrative tax records, and only some of the overstatement is due to the use of survey

[3] Wheaton and Stevens (2016) compared how these tax models, as well as Urban Institute's TRIM model and the Bakija (2014) model, affect estimates of the SPM in tax year 2012, but they did not have access to administrative tax records in their analysis. In general, TAXSIM relies on a file transfer protocol to feed in the input file from a remote server and send back the output file in a return exchange. This exchange is not suitable for use with our restricted data. In such cases, TAXSIM may be deployed using a local executable; however, while we had access to TAXSIM when using the internal version of the ASEC, we do not have permission to use it with IRS-provided variables.

income as inputs. However, when restricting the sample to a harmonized group of potential filers—dropping those with imputed earnings or no link to tax data and reweighting—we find that the survey tax simulators align closely with antipoverty estimates using actual EITC payments, with about 3.1 million persons lifted out of poverty by the EITC in 2016. The assumption of 100 percent take up in the simulators seems to balance out the seemingly incorrect payments in the actual EITC recipient file, yielding comparable antipoverty effects. Further analysis reveals that the efficacy of the survey tax simulators holds across the entire low-income distribution, and not just at the threshold delineating poverty status.

These results add to recent literature highlighting the bias on income distribution statistics from earnings nonresponse in the CPS ASEC (Hokayem, Bollinger, and Ziliak 2015; Bollinger, Hirsch, Hokayem, and Ziliak 2019). It is also complementary to important work on under-reporting of transfer program income and its effects on poverty alleviation (Meyer, Mok, and Sullivan 2015; Meyer and Mittag 2019).

II. Background on the EITC

The EITC was established in 1975 to incentivize work over welfare ("workfare") by providing a refundable tax credit to families with qualifying children and low earnings, thereby creating a subsidy to market wages. The credit has three ranges—the "phase-in" or subsidy range where the credit amount increases at a fixed rate as earnings increase; the "plateau" range where the maximum credit is attained and held fixed; and the "phase-out" range where the credit is tapered away as earnings increase. The EITC was initially modest in size, largely offsetting payroll tax liability; however, it was expanded in both generosity and reach with the Tax Reform Act of 1986 and subsequently with the Omnibus Budget Reconciliation Acts (OBRA) of 1990 and 1993. OBRA90 differentiated tax units into those with one qualifying child versus two or more, and provided a more generous credit to those with two or more children. OBRA93 further expanded access to single individuals with no dependents, and also substantially increased the generosity of the credit with qualifying children such that by 1996 the EITC nearly reached the fourth decile of the married-couple income distribution and 150 percent of the median for female-headed families (Ventry 2000). A temporary, higher subsidy tier was added for families with three or more qualifying children with the American Recovery and Reinvestment Act (ARRA) of 2009, which was then made permanent in 2015. Table 1 summarizes the key parameters of the EITC in tax years 2005 and 2016, coinciding with the start and end of our sample. Across tax years the maximum credit is adjusted upward with inflation, though 2016 also contains the new category with a maximum subsidy rate of 45 percent and credit of $6,269. From inception in 1975 through 2016 the number of EITC recipients grew four-fold to 27.4 million tax filers, with the average inflation-adjusted credit growing ten-fold to $2,437.[4]

[Table 1 here]

The growth in the credit has generated a voluminous research literature, including estimates of the antipoverty impact. While food stamps are estimated to be effective at reaching those in deep poverty (Tiehen, Jolliffe, and Smeeding 2015), the EITC is widely touted as the most successful program for low-income working families. As Hoynes and Patel (2018) note, the EITC can affect poverty mechanically via the credit amount on after-tax income, as well as behaviorally by affecting both the extensive and intensive margins of labor supply (Eissa and Leibman 1996; Meyer and Rosenbaum 2001; Neumark and Wascher 2001; Eissa and Hoynes 2004).[5] Ziliak (2015a), using public release versions of the CPS ASEC and the CPS tax model, estimated that the EITC lifted about 4 million persons out of poverty (based on the official

[4] Brookings/Urban Tax Policy Center http://www.taxpolicycenter.org/statistics/eitc-recipients

[5] Substantial labor supply response, but more recent paper by Kleven (2019) suggests no extensive margin response.

poverty line) per year in the decade prior to the Great Recession in 2008, and over 5 million at the peak of the recession. Hoynes and Patel, using the public CPS ASEC along with TAXSIM, estimated that the EITC lifted 3.4 million children in single-mother families out of poverty (the authors used the official poverty line but defined a broader resource definition that included some in-kind transfers and tax payments and credits). The Center on Budget and Policy Priorities (2016) estimated that 6.5 million people were lifted out of poverty in 2015 using the Census SPM and public CPS ASEC with the CPS tax model.[6]

A challenge facing research on the EITC is that none of the major household surveys in the U.S. collect information on credit receipt or amounts; more generally, they do not collect information on tax unit formation or tax deductions, and thus researchers either construct their own estimates of the EITC using program parameters or rely on publicly provided tax simulators such as the CPS tax model or TAXSIM (Ziliak 2015b). For the EITC the key inputs are earnings, interest income, age of filer, age and number of qualifying children, and filing status. The Census Bureau since the 1980s has used available survey information in the CPS ASEC to construct tax units and simulate federal, state, and payroll tax liability, including the EITC, and makes the output available in public release versions of the CPS ASEC. Each user of TAXSIM and the Bakija model, however, has to independently compile this information from the survey prior to submitting the input to either simulator. While most surveys have enough information to create pointers to family and household relationships, these may not translate directly into tax filing units; that is, a family or household may contain multiple filing units. In addition, determining who does and who does not meet the "qualifying child" test is difficult. Children under age 19

[6] The larger estimate in the Center on Budget and Policy Priorities report results from the fact that the SPM poverty threshold is higher up the income distribution than the official line, and thus can capture a larger share of the EITC recipient population.

must live with the filer at least six months and a day during the tax year, but the surveys do not typically record length of time in the household. This residency issue is further complicated by children of divorced parents who may spend equal time in each household. Children older than 18 years and younger than 24 may also be claimed as dependents if their primary activity is a full-time student, which is collected in some, but not all, household surveys. The net result is that the researcher must make assumptions on family relationships that may have direct impacts on the quality of tax estimates. The aim of our project is to use a direct match of survey to administrative tax records to assess how well these models perform relative to actual payment in the antipoverty effect of the EITC.

III. Data and EITC Simulators

The data we use in this project derive from a joint statistical agreement between the Census Bureau and the IRS. The survey data used are yearly internal CPS ASEC files from 2006 to 2017, linked at the individual level for the corresponding tax year with the IRS data (that is, 2005 to 2016).[7] The tax data included are Form 1040 individual income tax records, the EITC recipient file, the CP09/27 file (a record of taxpayers sent a notice from the IRS about their potential EITC eligibility), and Form W-2 wage and tax statements. The primary sample used is the same as that employed by the Census Bureau for its official poverty estimation, but restricted to those with a link to the tax data and who do not have their earnings imputed. We then reweight the data based on the probability that an observation was linked to tax data and did not have

[7] The internal CPS ASEC differs from the public-use CPS ASEC primarily in terms of top-codes of earnings and income. For example, the internal top code of earnings is $1.09 million, while it is $250,000 in public versions. Individuals in the CPS ASEC are linked to the tax data via a unique, protected identification key, or PIK. See Wagner and Layne (2014) for a description of the model.

imputed or edited information.[8] Details of sample construction and linking are provided in the Data Appendix.

With the linked CPS ASEC-IRS data, we consider six alternative approaches to estimating the EITC compared against the benchmark of actual EITC payment by the IRS, which we denote throughout as *IRS Paid*. There is some dispute whether administrative data should be treated as the "gold standard" as it too may suffer from measurement error (Kapteyn and Ypma 2007; Bollinger et al. 2019). However, in this case the administrative record is the correct benchmark because we wish to assess how well the other approaches align to the actual EITC dollars circulating in the economy, whether those dollars were correctly paid or not.

[Table 2 here]

Table 2 summarizes the alternative approaches to EITC estimation, and details are described in the Data Appendix. Briefly, we compare actual EITC payments against survey-based estimates of the EITC and hybrid, survey-administrative-data-based estimates of the EITC. The survey-based approaches use tax characteristics of families and their heads in the ASEC in conjunction with NBER's tax simulator (*TAXSIM*), Jon Bakija's simulator (*Bakija)*, and the Census Bureau's own internal estimate of taxes (*CPS*).[9] The *CPS* estimate of the EITC is provided by Census annually as part of the public release of the CPS ASEC, while we use the information in the CPS ASEC to construct our own tax units and taxable income for the *TAXSIM*

---

[8] Bollinger et al. (2019) show that earnings nonresponse in the CPS ASEC is over 30 percent on average and U-shaped across the earnings distribution—highest in left and right tails—and not missing at random, meaning that reweighting the data after dropping imputed values helps, but will not completely remove bias from imputation. Hokayem et al. (2015) show that imputation in the CPS ASEC leads to an official poverty rate that is too low by about 1 percentage point, and reweighting using an inverse probability weight fills about 70 percent of that gap.
[9] For simplicity, we refer to "CPS," "TAXSIM," "Bakija," and "Bakija admin," italicized, to refer to the estimated EITC payment derived from each simulator (with estimates greater than zero identifying someone as both eligible and paid, according to the model in question). The use of italics distinguishes the estimates from the simulators themselves. Any critique of these estimates should not be taken as critiques of the simulators, as the estimates are dependent on the quality of the input file.

and *Bakija* models. The hybrid survey-administrative data approaches include the Bakija simulator to estimate EITC payments, but in lieu of income information in the CPS ASEC, we use administrative income and earnings from Forms 1040 and W-2, which we denote as *Bakija Admin*. The remaining two approaches—denoted as *IRS Eligible* and *IRS Paid & Eligible*—are the same as those constructed as part of the joint Census-IRS project on estimating EITC take-up. The *IRS Eligible* model uses family structure information in the CPS ASEC, along with EITC policy parameters and income and earnings data from the 1040 and W-2, to estimate who is eligible for the EITC and how much they "should" receive if they were in fact to file. Thus, this model assumes 100% take-up like the survey-based simulators. The *IRS Paid & Eligible* model then restricts the *IRS Eligible* sample to those actually receiving the EITC via filing the 1040 (but still using simulated EITC payments). Comparing the EITC payments from the *IRS Eligible* and *IRS Paid & Eligible* samples reveals how much EITC money is potentially left unclaimed because of imperfect take-up, while the ratio of the number of recipients in the *IRS Paid & Eligible* sample to the *IRS Eligible* sample provides an estimate of EITC take-up.

Appendix Table 1 contains summary statistics from the sample, where the upper panel presents statistics at the individual tax filer level pooled across the 2005–2016 tax years, while the bottom panel is aggregated to the family level. All income amounts are in real 2009 dollars using the Personal Consumption Expenditure Deflator.[10]

IV. Alternative EITC Estimates Over Filer Characteristics

A first step in our process is an assessment of each of our six modeling estimates against the internal recipient file. Tables 3 and 4 compare each weighted EITC estimate to the internal

---

[10] Appendix Table B-3 in the 2017 Economic Report of the President, https://obamawhitehouse.archives.gov/sites/default/files/docs/appendix_b-statistical_tables_relating_to_income_employment_and_production_2017.pdf

EITC recipient file for tax years 2006 and 2016, respectively. Characteristics in the recipient file include filing status and number of children claimed for EITC, and are compared with the corresponding characteristic calculated using the *IRS Eligibility* estimation strategy. The survey-based estimates each cover between 67 and 68 percent of benefits in 2006 (between 70 and 71 in 2016) and between 75 and 85 percent of recipients (between 78 and 79 in 2016). Substituting administrative values of income and earnings (*Bakija Admin)* leads to slightly better coverage for benefits (72 percent in both 2006 and 2016), but better coverage for recipients only in the later year (75 percent in 2006 and 81 in 2016). *IRS Paid* in 2006 shows coverage of 95 percent of dollars and 92 percent of recipients, but only 79 percent and 81 percent coverage in 2016—a degradation in coverage rates that may be related to increasing non-response rates. The *IRS Eligible* estimate covers about 69 percent of dollars and 85 percent of recipients in 2006 (79 and 81 percent in 2016), while *IRS Eligible and Paid*, not unexpectedly, covers the smallest proportion: 59 percent of benefits and 66 percent of recipients in 2006 (63 percent and 73 percent in 2016).

[Tables 3 and 4 here]

While these results suggest that, overall, we capture a substantial portion of the paid population using the linked the data, Tables 3 and 4 show that the quality of the estimates vary by filer characteristic. For example, in the *IRS Paid* column we calculate between 10 (Table 4) and 34 (Table 3) percent more recipients with zero qualifying children than the internal file reports. This is due to persons who claim one or more children on their 1040 but do not appear to have a qualifying child according to the survey data. However, much of the heterogeneity arises from incorrect payments. Every estimate over-reports the dollars going to filers with zero children. While this category represents only three percent of total EITC expenditures from the recipient file, each estimate reports between 1.5 and six times more dollars accruing to this category. A look at the *IRS Paid & Eligible* ratio makes clear that many of those who are incorrectly paid have inappropriately claimed one or more children, which accounts for the discrepancy between the recipient file and the estimates. The discrepancy disappears once we limit the comparison only to those who are paid and eligible.

[Figures 1 and 2 here]

In Figures 1 and 2 we examine how well the modeling strategies translate into average credit amounts in each year, for the entire reweighted sample including zero credit amounts (Figure 1) and restricted to those deemed eligible by the *IRS Eligible* model. Figure 1 shows that the mean values from *TAXSIM*, *Bakija*, *CPS, Bakija Admin*, and *IRS Eligible* series are each quite close, increasing by about 25 percent during the Great Recession and credit expansion year of 2009, and gradually falling back to near pre-recession levels by 2016 as earnings increased. However, all the estimators fall short of the actual paid amount from the recipient file (*IRS Paid*) by $75-$100 depending on year.[11] Figure 2 shows the same series, but restricted to those deemed by the Census-IRS EITC project to be eligible for the credit, which accounts for the much higher average credit amounts. In most years, the survey-based estimators from *TAXSIM*, *Bakija*, and *CPS* are very similar to each other and to actual payments with *IRS Paid*, at least after the recession. Notably, because we condition on eligibility, the gap of approximately $250 between *IRS Eligible* and *IRS Paid* reflects the difference in average credit value induced by imperfect take-up—those eligible but failing to take up benefits would receive more on average.

[Table 5 here]

[11] Appendix Table A1 shows that across all sample years credit amounts, inclusive of zeros, average $248, $246, and $245, respectively, for *TAXSIM*, *Bakija*, and *CPS*, while *IRS Paid* averages $331, which is larger than would be expected based on official eligibility criteria, i.e. the average credit for *IRS Eligible* is $250.

This leads us to the question of who appears eligible, who takes up, and who appears to take up incorrectly in the official estimates. Table 5 provides the breakdown of these categories and reports the mean values of income and demographic characteristics. On average, conditional on eligibility, an eligible receiver is paid $2,270 in credit but looks eligible for slightly less ($1,945); their survey-based credit value is even lower, at $1,586. Non-eligible receivers are paid close to the same credit amount on average and have considerably lower mean survey-based amounts ($239). Meanwhile, eligible non-receivers pass up, on average, about $1,082 ($807 according to our survey estimate). Eligible non-receivers have the lowest IRS annual earnings and income, reflecting lower take-up rates in the phase-in region of the credit compared with the plateau or phase-out range. While ineligible receivers have higher income and earnings than eligible receivers, they are poorer compared with ineligible non-receivers. Non-eligible receivers are also more likely to be married and have fewer children than eligible receivers.

Distributional Differences

We move beyond mean outcomes to examine the value of estimated EITC payments across the income distribution. For each estimate, we calculate the average value of the credit over bins of individual income, where the income measure is that used in estimating the credit value. In other words, *IRS Paid* and *IRS Eligible* are averaged over bins of IRS 1040 AGI, while *TAXSIM*, *Bakija* and *CPS* are averaged over bins of CPS ASEC survey individual income.[12]

[Figure 3 here]

We graph the estimates using survey data alone in the top panel of Figure 3. All three simulator estimates are very close to one another and trace out the general shape of the EITC schedule. There is a drop to zero in credit receipt at the point in the income distribution where

[12] Recall that while earnings are necessary to qualify for the EITC, actual eligibility and credit amounts are determined by both earnings and AGI.

one would expect to see it (around $45,000). We next gray out these estimates, in the bottom panel, and trace out the IRS-derived estimates, including *IRS Paid*, *IRS Eligible*, and *Bakija Admin*. This panel shows that the survey-based estimates in gray are similar over the distribution compared to hybrid *IRS Eligible* and *Bakija Admin* estimates, though the latter lie everywhere below the survey-based estimates in the $10,000 - $20,000 range, which as we will see below has ramifications for anti-poverty effects. Foreshadowed by the higher average estimate in Figure 1, *IRS Paid* lies above all other estimates over much of the income distribution. A clear finding of this exercise is that the difference between *IRS Paid* and *IRS Eligible* is widest at low levels of individual income, indicating that the payment of EITC is targeted to low-income people regardless of whether they are strictly eligible according to other parameters.[13]

[Figures 4 and 5 here]

Figure 4 breaks out the EITC distributions by number of children claimed and Figure 5 by filing status. These variables combine information from both the survey and IRS data to generate the most likely claiming status for each tax unit. For ease of presentation, and because of the comparability of estimates, in this set of figures *TAXSIM* stands in for all of the survey-based estimators. Our choice of *TAXSIM* as a stand-in derives from its easy accessibility and its widespread use among researchers. While the survey estimators come close to both *IRS Eligible* and *IRS Paid* for tax units with one or more children and for joint filers, the same cannot be said for units without children and with single filers. By definition, unmarried filers with children are modeled as heads of household. Thus, due to incorrect filing by single filers without children in the household, *IRS Paid* lies significantly above the survey estimates for single filers. Nevertheless, based on the scale of the y-axis, the dollar amounts that apparently are claimed but

[13] This is not to deny that there are clearly tax filers beyond the $45,000 mark receiving credit dollars, apparently incorrectly.

not owed do not account for much of overall EITC expenditure. Our estimate using *Bakija Admin* lies below the survey-based eligibility estimates in the $10,000 - $20,000 range when children are present, and when filing status is either head of household or married. Bollinger et al. (2019) find that earnings in administrative data lie below those in the CPS ASEC at the lower tail of the distribution, perhaps because some earnings are not taxed or are not reported. This is consistent with EITC amounts being higher with survey income data than with tax income data underlying the *Bakija Admin* credit in Figures 4 and 5.

V. Alternative EITC Estimates Over the Family Income Distribution

The EITC is targeted at the tax filing unit, but the official measure of poverty in the United States is based on resources available to the family, which the Census Bureau defines as two or more persons related by birth, adoption, or marriage. Related subfamilies are assigned the poverty status of the main family, while unrelated individuals retain the poverty status determined by their own resources.[14] However, the EITC can be received by both the main family and the related subfamily provided that they each file independent tax returns. Thus, in order to assess the anti-poverty effects of the EITC we need to move from incomes of the tax filing unit to incomes of the family. Previously, we restricted our analysis to a sample of CPS ASEC respondents who reported earnings and income and who received the unique identifier (called the person identification key, or PIK) used in matching to administrative records. Those who do not respond to the CPS ASEC earnings and income questions have their responses imputed using a hot-deck procedure.[15] To produce estimates that align with Census poverty definitions, we must expand the sample to all families, including those who have imputed responses or who lack an administrative record.

[14] Importantly, though, cohabiting partners resources are not considered in official poverty measurement.
[15] See Welniak (1990) and Bollinger et al. (2019) for description of the Census hot-deck imputation procedure.

Figure 6 provides some evidence on the challenges this choice presents in terms of accurate estimates. We define four groups based on their response categories: respondents with a PIK, who have both reported survey values and potential IRS values (69 percent of family heads); nonrespondents with a PIK, who have imputed survey values but potential IRS values (20 percent); respondents without a PIK, who only have reported survey values (6 percent); and nonrespondents imputes without a PIK, whose EITC estimates derive solely from imputed survey values (5 percent).

[Figure 6 here]

For the first group, credit values based on survey-reported earnings and income are consistently higher over the income distribution than when using administrative records for both the income bins and the credit calculation, which is consistent with Figures 3-5 that are restricted to tax-unit respondents with a PIK. While this is also true for the second group, the two distribution lines are closer together, stemming more from the *Bakija Admin* shifting up to be more aligned with *TAXSIM*. This is consistent with higher nonresponse at the left tail of the distribution (Hokayem et al. 205; Bollinger et al. 2019). Hokeyem et al. report that the hotdeck assigns too high of earnings, which is why official poverty is understated. Meanwhile, those who respond but do not have a corresponding administrative record appear to have similar credit values over the income distribution to other survey respondents, though with a higher peak as non-PIK families tend to be lower income and thus eligible for a higher credit (Bollinger et al. 2019). The nonrespondent and non-PIK group are clearly different, with credit values considerably lower over the income distribution than calculated for the other groups.

Moving forward, our procedure is to examine family-based calculations of credit receipt based on the full sample and survey responses, whether self-reported or imputed. An estimate from TAXSIM stands in for all survey-based estimates. We then compare this to an administrative-data estimate from the Bakija procedure and, finally, the *IRS Paid* value from a direct match to the EITC recipient file. The latter two estimates will capture the approximately 90 percent of CPS ASEC family heads who are assigned a PIK. As a final comparison, we then return to a family-based version of the restricted respondent and PIK sample, which will allow us to include the *IRS Eligible* and *IRS Paid and Eligible* estimates.

[Figures 7 and 8 here]

In Figure 7, we depict *TAXSIM*, *IRS Paid*, and *Bakija Admin* in bins of family income from the CPS ASEC in the top panel and by income-to-needs in the bottom panel.[16] Here, we see that from the family-income perspective there is a substantial right tail in *IRS Paid* compared to the survey-based and hybrid estimators of the EITC. Figure 7 suggests that actual EITC payments are not as target efficient at the family level compared to the individual tax unit level as depicted in the prior figures. This is not necessarily leakage in the usual sense because the tax unit and the family unit are not the same concepts, and that families may contain multiple tax filers. The family income graphs will pick up these multi-filer units as demonstrated in Figure 8. It is clear from Figure 8 that the density in the multi-filer panel is thicker at higher income levels, and does not taper off after $40,000 in family income as in the single filer case.[17] However, even in the single filer case we see that there is a thicker right tail in *IRS Paid* at higher income amounts, which may in fact reflect program leakage.

VI. Anti-Poverty Effects of the EITC

[16] Income-to-needs is computed by taking the ratio of family income to family-size specific poverty thresholds. Income-to-needs less than 1 indicates living in poverty.

[17] Jones and O'Hara (2017) show that multifiler households exhibit some strategic behavior in claiming EITC, using the movement of children across filers to increase EITC receipt.

We next examine how the alternative modeling strategies align in terms of the number of persons lifted above the family-size specific official poverty threshold each tax year by the EITC. Here we conduct two exercises. First, we examine the "off the shelf" type of calculation by the typical user of the full-sample CPS ASEC and either the TAXSIM, Bakija, or CPS tax models, where the full CPS ASEC sample includes persons with imputed earnings and incomes and with and without a link to tax data. We then compare these estimates directly to our benchmark value, which is *IRS Paid* in the full sample, meaning with imputed incomes but the actual EITC tax value only occurs with those tax units who can be linked (about 90% in a typical year). Second, we restrict attention to the original sample used in the earlier analysis in order to conduct more of an "apples-to-apples" comparison of the six modeling strategies. It is only the latter exercise where we are able to report the *IRS Eligible* and *IRS Paid & Eligible* series. These two exercises are informative for different purposes: the first allows us to use the official definition of poverty calculated from the full CPS ASEC and examine the impact of actual EITC receipt on that rate; the second allows us to compare estimation strategies over a specific set of potential filers.

[Figure 9 here]

The top panel of Figure 9 depicts the total number of persons lifted out of poverty by the EITC after adding in the EITC from *TAXSIM*, *Bakija*, *CPS*, and *IRS Paid*, and the bottom panel contains the number of children lifted out of poverty.[18] The three survey-based tax simulators predict that the EITC lifts 4.8 million persons out of poverty in an average year (about 2.75

[18] Appendix Table A3 contains the estimated poverty rates, where the top panel A is for the full CPS ASEC and the bottom panel B is for the reweighted restricted sample. The column labeled CPS ASEC in panel A replicates the official poverty rate as reported in the annual P60 report (Proctor et al. 2016), while the remaining four columns contain the corresponding estimated poverty rate after adding in the EITC from *CPS*, *TAXSIM*, *Bakija*, and *IRS Paid*.

million children in the bottom panel), while *IRS Paid* lifts 3.2 million persons in an average year, or 33 percent fewer, and this gap was exacerbated over the half decade after the Great Recession. Consistent with the declining average EITC payment in Figure 1, the antipoverty effects of the EITC fell towards pre-recession levels after 2014.

[Figure 10 here]

In Figure 10 we return to our original sample where we restrict the analysis to the sample without imputations and with a tax link, and reweight. We now find that all the survey-based tax simulators, as well as the hybrid survey and administrative *IRS Eligible* model, align very closely with *IRS Paid* in lifting about 3.3 million persons out of poverty. The difference between Figures 9 and 10 resides with the inclusion of imputed earnings in the full CPS ASEC sample of Figure 9. The Census imputation procedure assigns earnings higher up the distribution such that when income from the EITC is included, too many persons are artificially lifted out of poverty. The *Bakija Admin* series, however, is too low in its EITC antipoverty effects as anticipated because the income reported to the IRS is below that reported in the CPS ASEC in the lower tail of the distribution.

To gain a deeper look at where in the distribution the EITC is affecting family poverty, we present TIP curves (Jenkins and Lambert 1997), which represents graphically the incidence, intensity, and inequality of poverty (the three "I"s of poverty, or TIP). The incidence of poverty is simply the standard poverty rate, the intensity of poverty is the dollar amount necessary to lift people out of poverty as measured by the (normalized) poverty gap, and the inequality of poverty is the distribution of income among the poor. The TIP curve as such is defined as the sum of poverty gaps or normalized poverty gaps; that is, let the poverty threshold be represented by z and family income as y, then the poverty gap for family q is denoted as

(1) $$g_q = \max\{z - y_q, 0\}$$

and the normalized gap as

(2) $$\Gamma_y = \max\{\frac{z - y_q}{z}, 0\}.$$

Then for the cumulative population share p, $0 \le p \le 1$, the TIP curve $TIP(g;p)$ takes a value of 0 when $p = 0$, and takes a value of

(3) $$TIP(g;\frac{Q}{n}) = \frac{1}{n}\sum_{q=1}^{Q} g_q$$

for integer values of $Q \le n$. When $Q = n$, $TIP(g;1)$ is simply equal to the average poverty gap in the population. For intermediate values of p, $TIP(g;p)$ is determined by linear interpolation. The TIP curve for normalized gaps, $TIP(\Gamma;p)$, is found in an identical fashion but with the normalized gaps in (2) replacing the non-normalized gaps in (1). Thus, for each p, $TIP(g;p)$ and $TIP(\Gamma;p)$ are indexes of poverty for the 100*p*% poorest.

[Figure 11 here]

In Figure 11 we present the TIP curves for 2005, 2010, and 2015 using the official poverty threshold (*z*) in the top panel and 125%**z* in the bottom panel. The horizontal axis measures the cumulative population share, *p*, while the vertical axis measures the cumulative sum of per capita poverty gaps. The poverty rate is that *p* where the TIP curve becomes horizontal. The height of the TIP curve reflects the intensity of poverty, which in the case of $TIP(g;p)$ is the poverty gap per income-receiving family and/or individual, and the curvature of the TIP curve represents the inequality of poverty. A line from the origin to the point where the TIP curve becomes horizontal reflects equality of income among the poor. The TIP curve, then, is akin to an inverse Lorenz curve but where the unit of analysis is poverty gaps rather than income distributions. The figure makes it unambiguously clear that the incidence and intensity of poverty increased substantially with the Great Recession, but the EITC played a strong role in reducing the intensity of poverty, lowering per capita poverty gaps by about $200, as well as the inequality of poverty given the flattening out of the TIP curve vis-à-vis the curve with no EITC credit. The tax simulators tend to capture well the antipoverty effectiveness of the EITC across the distribution of income among the poor, differing no more than by $10-$20 in per capita gaps.

VI. Conclusion

We used unique linked survey and administrative tax data to assess the coverage and antipoverty effects of the EITC using survey-based tax simulators in comparison to actual EITC payments from the IRS. The simulation programs utilize researcher-provided model-based assumptions on who is and who is not eligible for the EITC based on survey values, and conditional on eligibility, assume that participation is 100 percent. Claims regarding the impact of the program on measures of poverty may be overstated if the credit estimation relies solely on survey data and if take-up and incorrect payment are not accounted for.

We find that, when using the full CPS ASEC with imputed earnings and income, EITC benefits estimated from the survey-based tax modules and their impact on poverty are overestimated compared with actual benefit payments. However, removal of observations with Census-imputed earnings and incomes results in comparable antipoverty estimates using commonly deployed tax simulators in survey data against administrative records. We take this as good news for the research community who generally only have access to publicly released versions of survey data and the tax simulators. We also find that actual EITC payments, even if incorrectly paid, are target efficient at the tax unit level. We do find some leakage at higher levels of the family income distribution, but this is mostly accounted for by multi-tax filing unit families. The EITC is shown to be effective at reducing not only the number of persons in poverty, but also the intensity and inequality of poverty, and thus playing a key redistributive role across the low-income population. Our takeaway for practitioners, consistent with the work of Hokayem et al. (2015) and Bollinger et al. (2019), is that users of the CPS ASEC drop observations with imputed earnings and at a minimum reweight the sample when conducting distributional research.

References

Bakija, Jon. 2009. "Documentation for a Comprehensive Historical US Federal and State Income Tax Calculator Program." Williamstown, MA: Williams College.

Bollinger, Christopher R., Barry Hirsch, Charles Hokayem, and James P. Ziliak. 2019. "Trouble in the Tails? What We Know About Earnings Nonresponse Thirty Years After Lillard, Smith, and Welch." *Journal of Political Economy*, 129(5): 2143-2185.

Center on Budget and Policy Priorities. 2016. "Policy Basics: The Earned Income Tax Credit."

Chetty, Raj, John Freidman, and Emmanuel Saez. 2013. "Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings," *The American Economic Review* 103(7): 2683-2721.

Eissa, Nada, and Hilary Hoynes. 2004. "Taxes and the Labor Market Participation of Married Couples: The Earned Income Tax Credit." *Journal of Public Economics* 88(9-10): 1931-1958.

Eissa, Nada, and Jeffrey Leibman. 1996. "Labor Supply Response to the Earned Income Tax Credit." *The Quarterly Journal of Economics* 111(2): 605-637.

Feenberg, Daniel, and Elisabeth Coutts. 1993. "An Introduction to the TAXSIM Model," *Journal of Policy Analysis and Management*, 12(1): 189-194.

Hokayem, Charles, Christopher R. Bollinger, and James P. Ziliak. 2015. "The Role of CPS Nonresponse in the Measurement of Poverty." *Journal of the American Statistical Association* 110(511): 935-945.

Hotz, V. Joseph, and John Karl Scholz. 2003. "The Earned Income Tax Credit," In *Means-Tested Transfer Programs in the United States*, R.A. Moffitt (ed), Chicago: NBER and University of Chicago Press, 141-197.

Hoynes, Hilary, and Ankur Patel. 2018. "Effective Policy for Reducing Poverty and Inequality? The Earned Income Tax Credit and the Distribution of Income," *Journal of Human Resources*, 53(4): 859-890.

Jenkins, S. and P. Lambert. 1997. "Three 'I's' of Poverty Curves, with an Analysis of UK Poverty Trends." *Oxford Economic Papers* 49: 317–327.

Jones, Maggie R. 2014. "Changes in EITC Eligibility and Participation, 2005-2009," U.S. Census Bureau, Center for Administrative Records Research and Applications, CARRA-WP-2014-04.

Jones, Maggie R. 2018. Tax Preparers, Refund Anticipation Products, and EITC Noncompliance. Unpublished manuscript.

Jones, Maggie R., and Amy B. O'Hara. 2016. "Do doubled-up families minimize household-level tax burden?" *National Tax Journal* 69(3): 613-640.

Kapteyn, Arie and Jelmer Y. Ypma. 2007. "Measurement Error and Misclassification: A Comparison of Survey and Administrative Data." *Journal of Labor Economics* 25(3): 513-551.

Marcuss, Rosemary, Alain Dubois, Janice Hedemann, Mary-Helen Risler, and Kara Leibel. 2014. *Compliance Estimates for the Earned Income Tax Credit Claimed on 2006-2008 Returns*. Internal Revenue Service, Publication 5162, Washington DC.

Meyer, Bruce D. 2010. "The Effects of the Earned Income Tax Credit and Recent Reforms," In *Tax Policy and the Economy*, 24(1): 153-180.

Meyer, Bruce and Dan Rosenbaum. 2001. "Welfare, the Earned Income Tax Credit, and the Labor Supply of Single Mothers." *Quarterly Journal of Economics* 116(3): 1063-1114.

Meyer, Bruce, and Nikolas Mittag. 2019. "Using Linked Survey and Administrative Data to Better Measure Income: Implications for Poverty, Program Effectiveness, and Holes in the Safety Net," *American Economic Journal: Applied Economics* 11(2): 176-204.

Meyer, Bruce, Wallace Mok, and James Sullivan. 2015. "Household Surveys in Crisis," *Journal of Economic Perspectives* 29(4): 199-226.

Neumark, David, and William Wascher. 2001. "Using the EITC to Help Poor Families: New Evidence and a Comparison with the Minimum Wage," *National Tax Journal* 54(2): 281-317.

Nichols, Austin, and Jesse Rothstein. 2016. "The Earned Income Tax Credit," In *Economics of Means-Tested Transfer Programs in the United States, Volume 1*, Robert A. Moffitt (ed.), Chicago, IL: The University of Chicago Press, 137-218.

O'Hara, Amy. 2004. "New Methods for Simulating CPS Taxes," U.S. Census Bureau, SEHSD Working Paper 2004-08.

Plueger, Dean. 2009. "Earned Income Tax Credit Participation Rate for Tax Year 2015," Internal Revenue Service Working Paper, https://www.irs.gov/pub/irs-soi/09resconeitcpart.pdf.

Proctor, Bernadette D., Jessica L. Semega, and Melissa A. Kollar. 2016. "Income and Poverty in the United States," U.S. Census Bureau, Current Population Reports, P60-256.

Renwick, Trudi, and Liana Fox. 2016. "The Supplemental Poverty Measure: 2015," U.S. Census Bureau, Current Population Reports, P60-258.

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

Tiehen, Laura, Dean Jolliffe, and Timothy Smeeding. 2015. "The Effect of SNAP on Poverty," In *SNAP Matters: How Food Stamps Affect Health and Well-Being*, J. Bartfeld, C. Gundersen, T. Smeeding, and J. Ziliak (eds), Stanford: Stanford University Press, 49-73.

U.S. Bureau of the Census. 1993. *Measuring the Effect of Benefits and Taxes on Income and Poverty: 1992*, Current Population Reports, Series P60-186RD, U.S. Government Printing Office, Washington, DC.

Ventry, Dennis. 2000. "The Collision of Tax and Welfare Politics: The Political History of the Earned Income Tax Credit, 1969-99". *National Tax Journal* 53(4, Part 2): 983–1026.

Wagner, Deborah, and Mary Layne. 2014. "The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications' (CARRA) Record Linkage Software," CARRA Working Paper Series #2014-01, U.S. Census Bureau.

Welniak, Edward J. 1990. "Effects of the March Current Population Survey's New Processing System on Estimates of Income and Poverty." US Census Bureau, Washington, DC.

Wheaton, Laura, and Kathryn Stevens. 2016. "The Effect of Different Tax Calculators on the Supplemental Poverty Measure," The Urban Institute.

Ziliak, James P. 2015a. "Recent Developments in Antipoverty Policies in the United States." In *Social Policies in an Age of Austerity*, J.K. Scholz, H. Moon, and S. Lee (editors), Cheltenham, UK: Edward Elgar Publishing, 235-262.

Ziliak, James P. 2015b. "Income, Program Participation, Poverty, and Financial Vulnerability: Research and Data Needs," *Journal of Economic and Social Measurement* 40(1-4): 27-68.

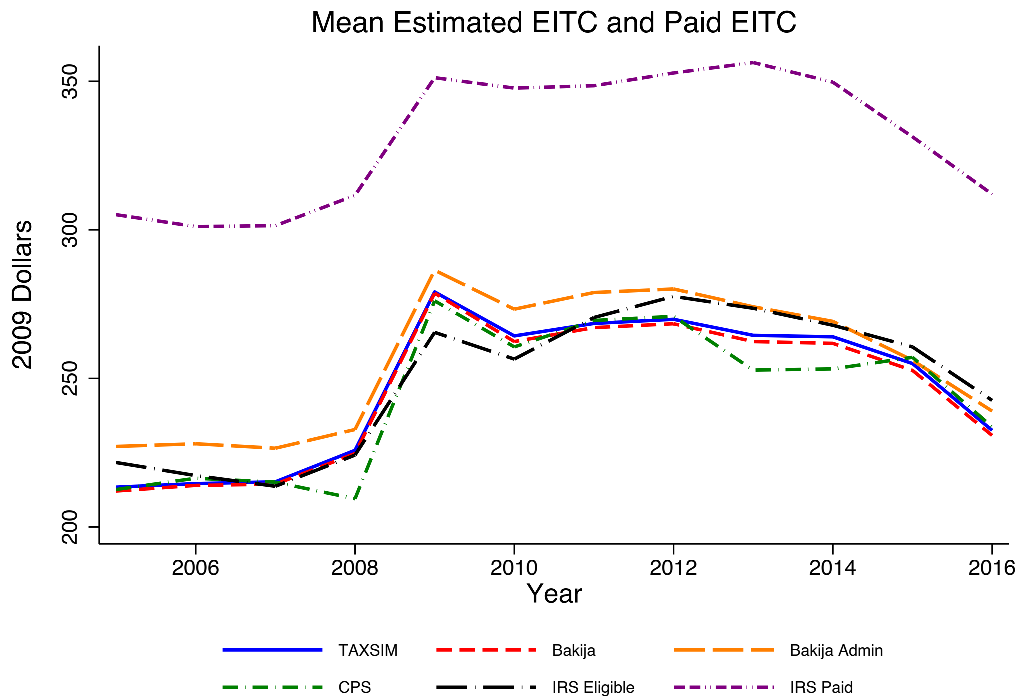


Figure 1. Mean estimated EITC from TAXSIM, Bakija, and Census CPS tax models, and estimated and paid EITC from combined CPS ASEC and IRS data. Source: combined EITC/CP0927 recipient files, CPS ASEC, Form 1040, and Form W-2, tax years 2005-2016. Release authorization number CBDRB-FY20-224.

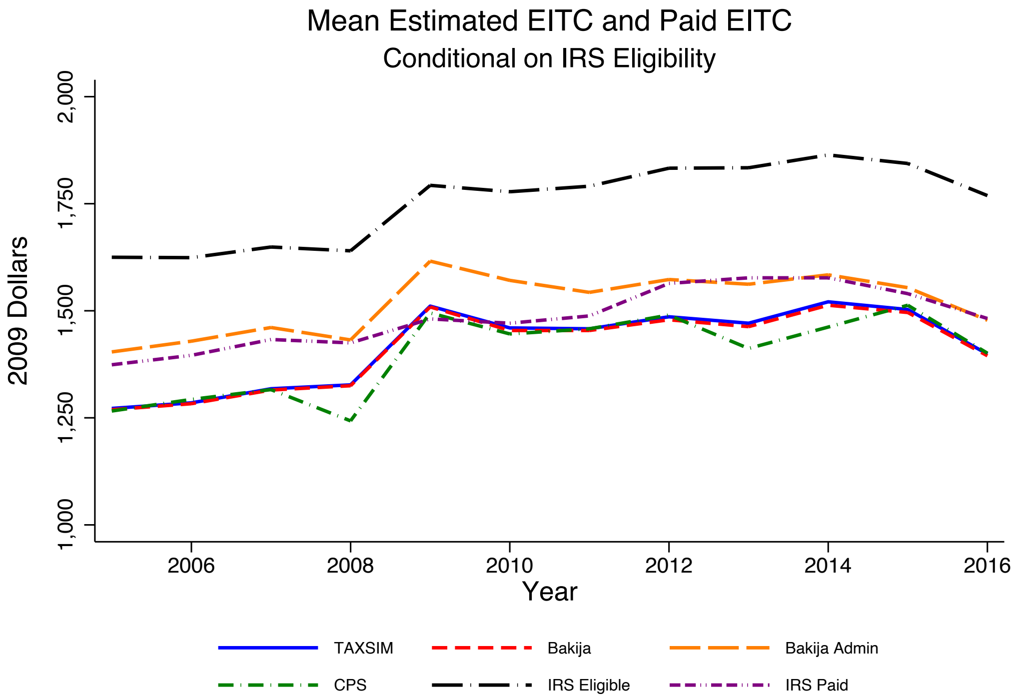
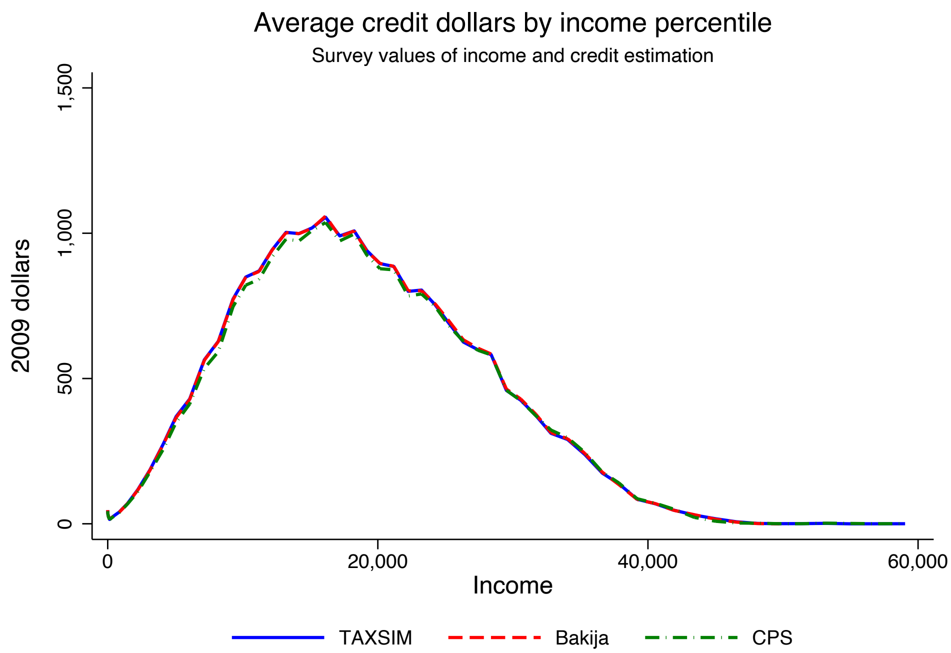


Figure 2. Mean estimated EITC from TAXSIM, Bakija, and Census CPS tax models, and estimated and paid EITC from combined CPS ASEC and IRS data. Source: combined EITC/CP0927 recipient files, ASEC, Form 1040, and Form W-2, tax years 2005-2016. Release authorization number CBDRB-FY20-224.



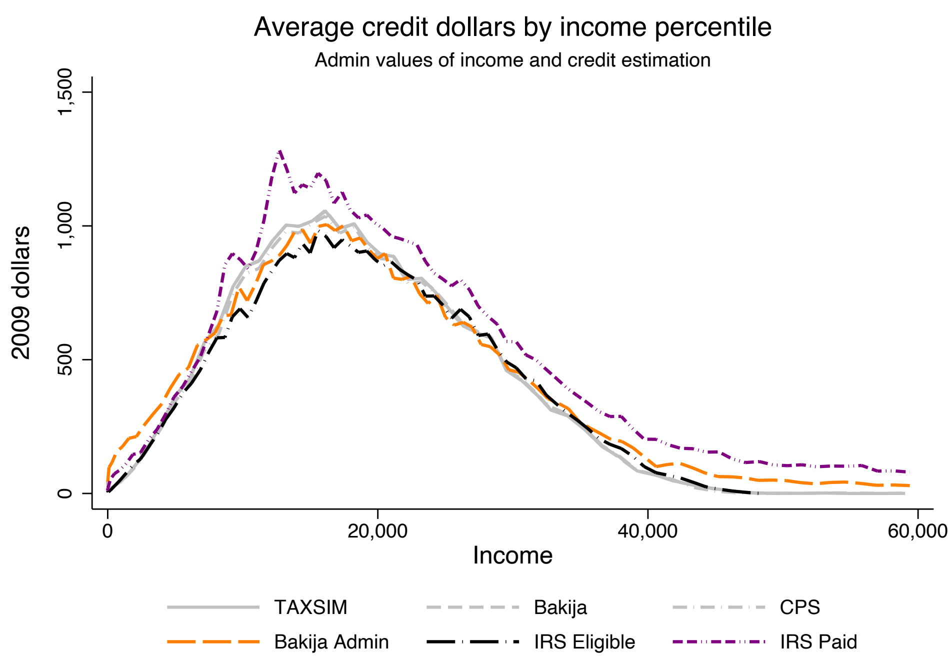


Figure 3. Top: Average credit dollars by income percentile for estimates using CPS-derived values of income. Bottom: Values of estimates derived from using administrative tax records for both income and EITC receipt. Source: combined EITC/CP0927 recipient files, CPS ASEC, Form 1040, and Form W-2, tax years 2005-2016. Release authorization number CBDRB-FY20-224.

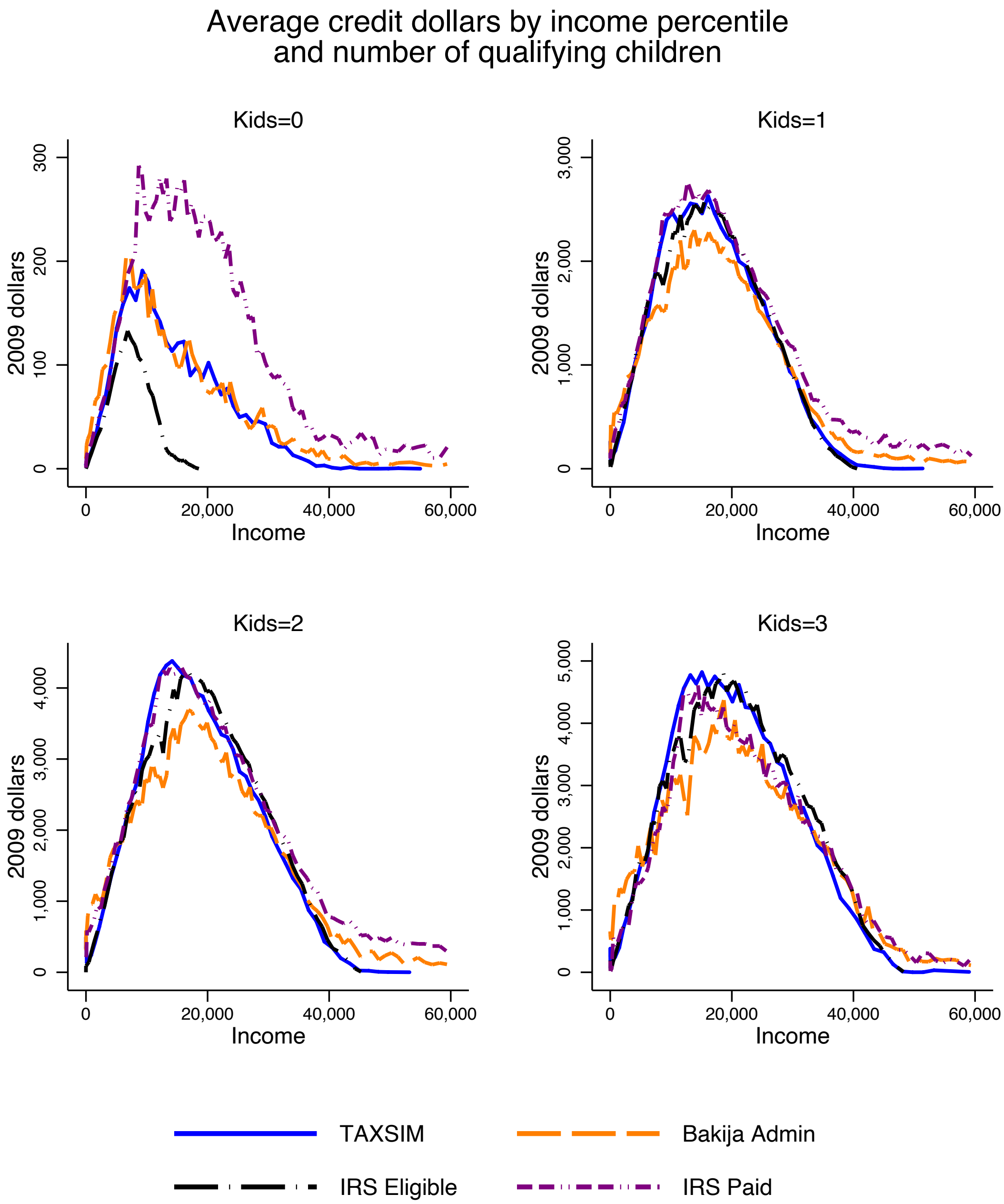


Figure 4. Average credit dollars by bins of income, broken out by number of children in household. Source: combined EITC/CP0927 recipient files, CPS ASEC, Form 1040, and Form W-2, tax year 2005-2016. Release authorization number CBDRB-FY20-224.

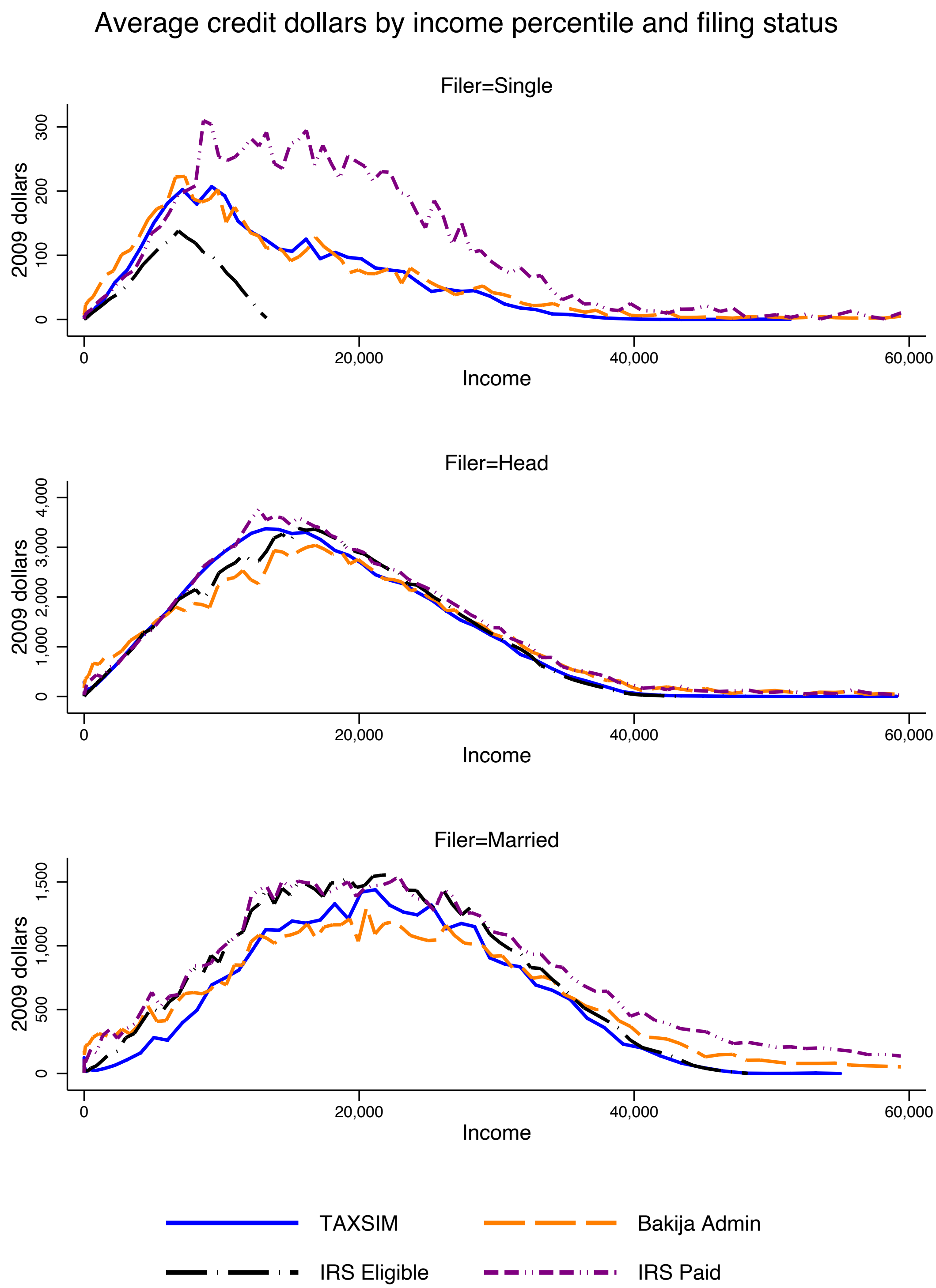


Figure 5. Average credit dollars by bins of income, broken out by filing status of tax unit. Source: combined EITC/CP0927 recipient files, CPS ASEC, Form 1040, and Form W-2, tax years 2005-2016. Release authorization number CBDRB-FY20-224.

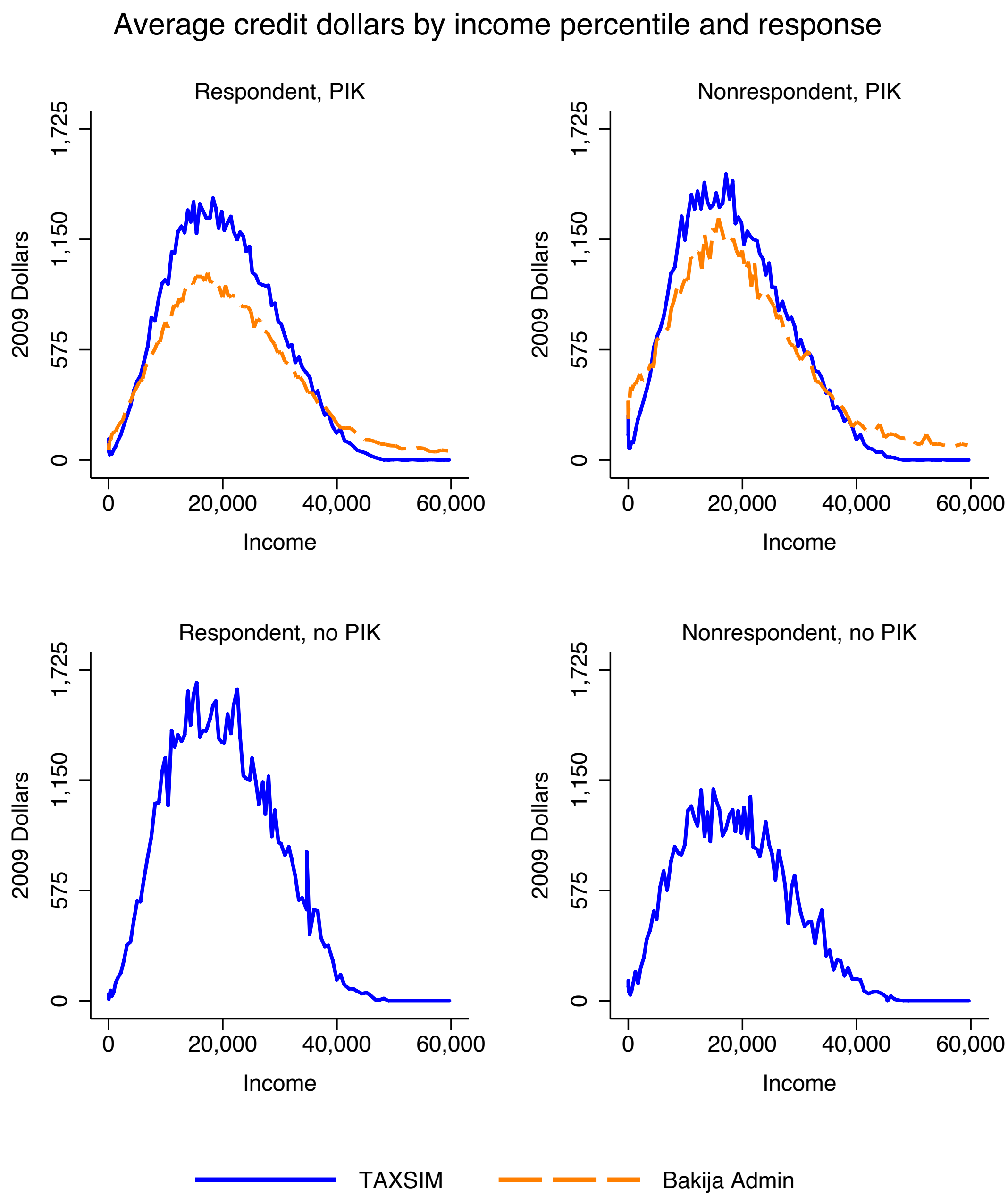
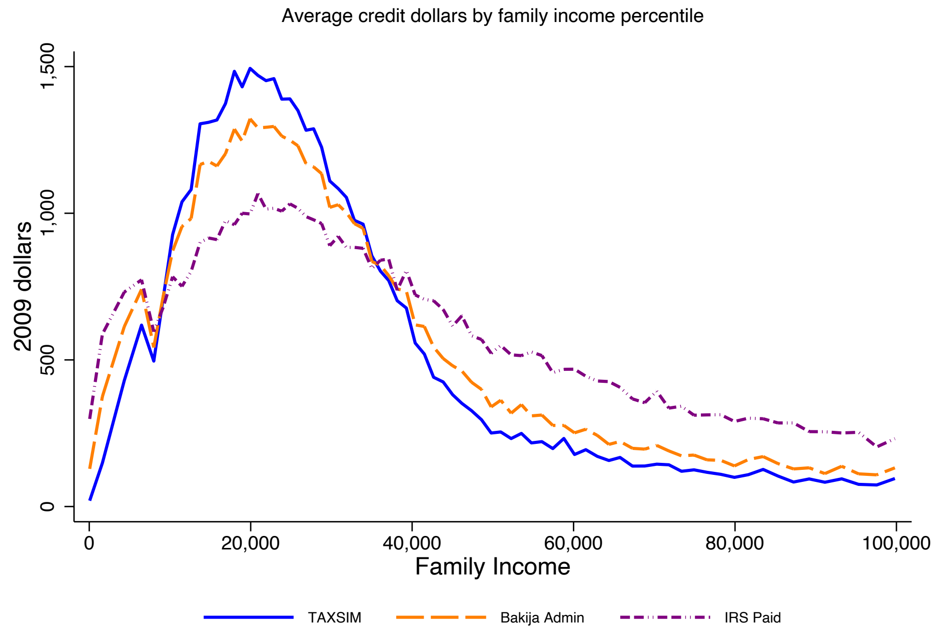


Figure 6. Survey- and administrative-records-based estimates of EITC by response and identifier placement (PIK) category. Source: combined EITC/CP0927 recipient files, CPS ASEC, Form 1040, and Form W-2, tax year 2005-2016. Release authorization number CBDRB-FY20-271.



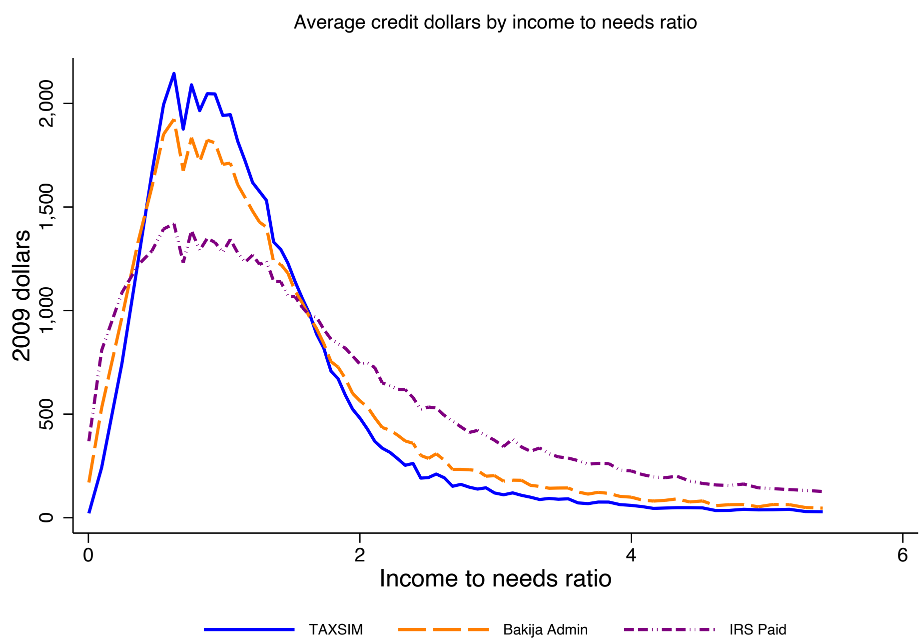


Figure 7. Average credit dollars by bins of family income. Here, the *TAXSIM* estimate stands in for all survey-based estimates, as these were closely aligned. Full poverty sample. Source: combined EITC/CP0927 recipient files, CPS ASEC, Form 1040, and Form W-2, tax years 2005-2016. Release authorization number CBDRB-FY20-224.

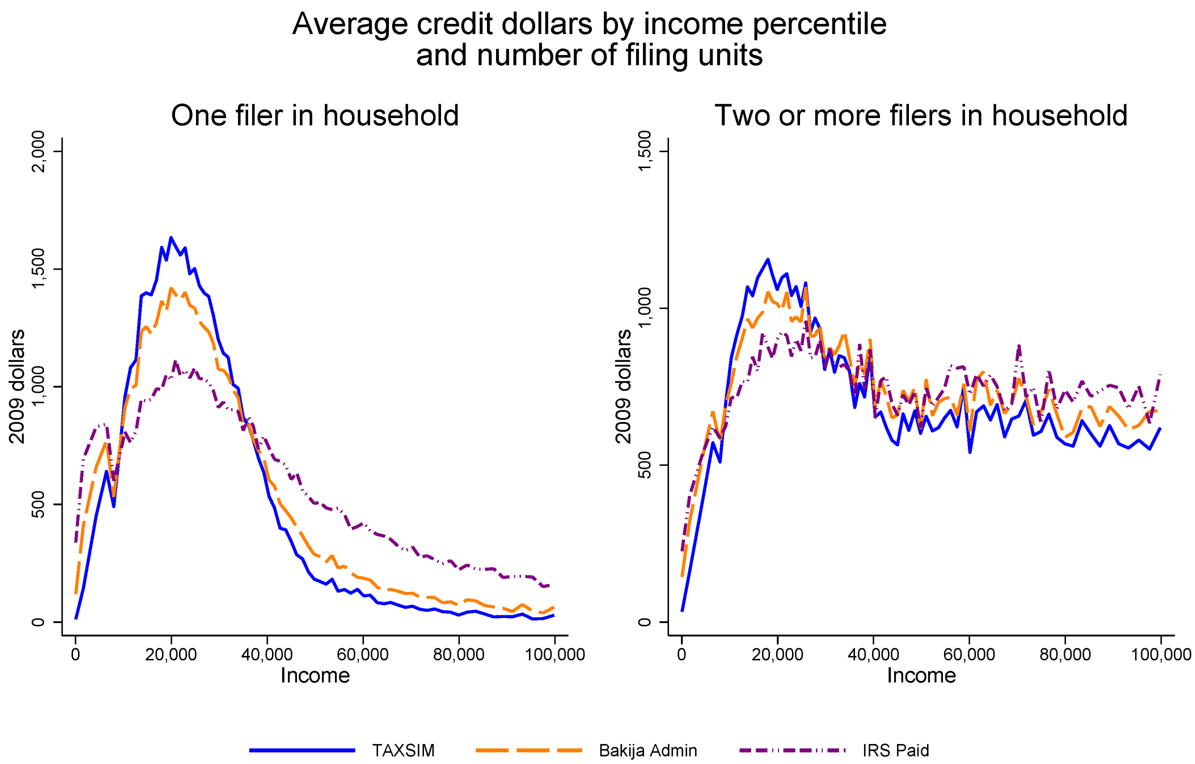
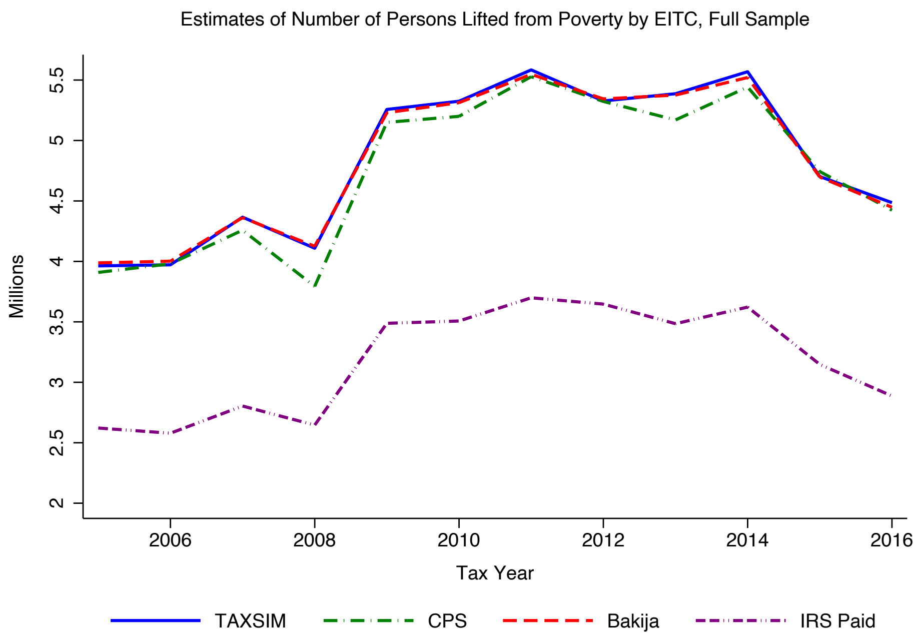


Figure 8. Average credit dollars by bins of family income, by number of tax filing units. Here, the *TAXSIM* estimate stands in for all survey-based estimates, as these were closely aligned. Full poverty sample. Source: combined EITC/CP0927 recipient files, CPS ASEC, Form 1040, and Form W-2, tax year 2005-2016. Release authorization number CBDRB-FY20-224.

A. All Persons



B. Children

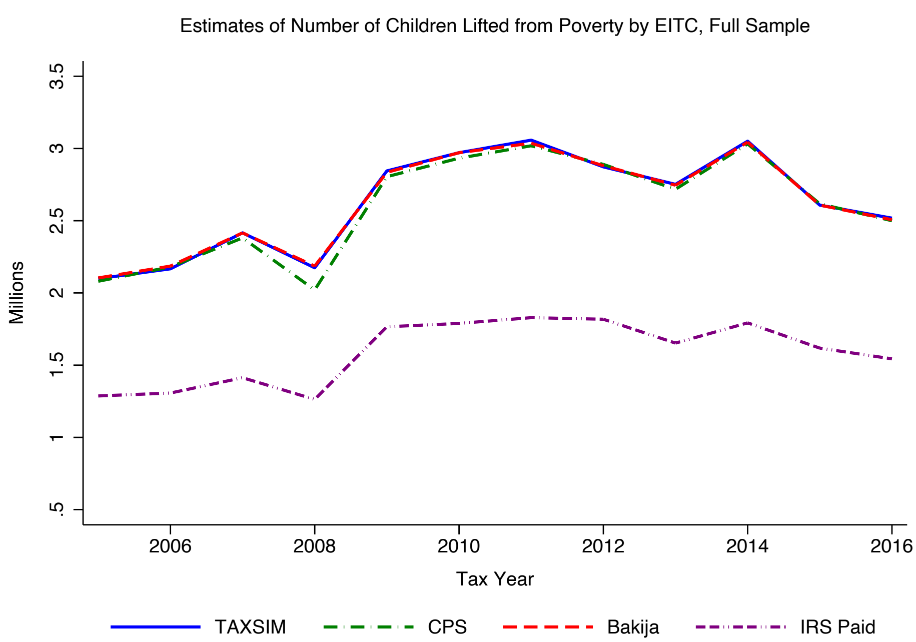


Figure 9. Estimates of the number of persons lifted from poverty by EITC using the full sample and original survey weights. Source: combined EITC/CP0927 recipient files, CPS ASEC, Form 1040, and Form W-2, tax year 2005-2016. Release authorization number CBDRB-FY20-224.

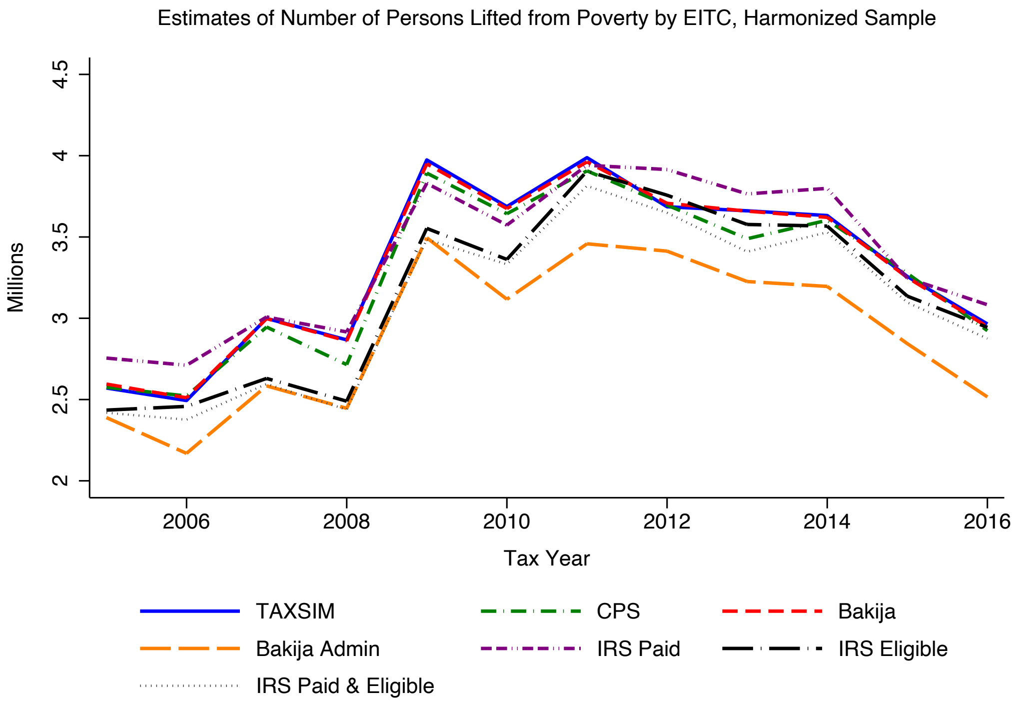
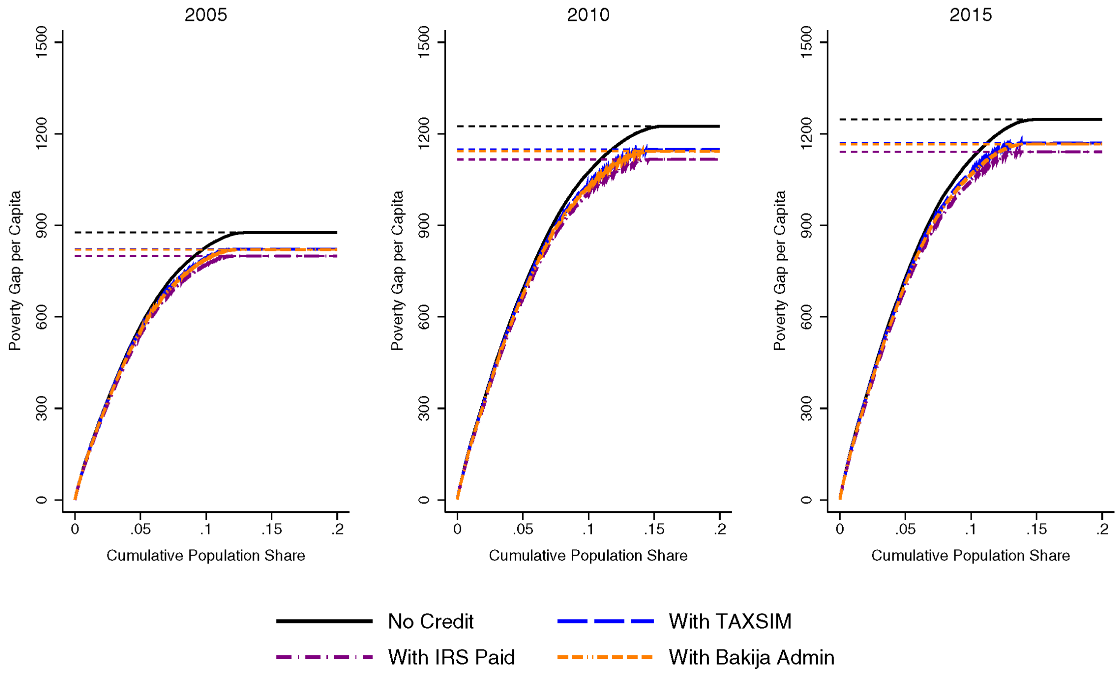


Figure 10. Estimates of the number of persons lifted from poverty by EITC using the sample and weights adjusted for link between the ASEC and IRS data and response to earnings and income questions in the ASEC. Source: combined EITC/CP0927 recipient files, CPS ASEC, Form 1040, and Form W-2, tax years 2005-2016. Release authorization number CBDRB-FY20-224.

A. Official Poverty Threshold



B. 125% of Official Poverty Threshold

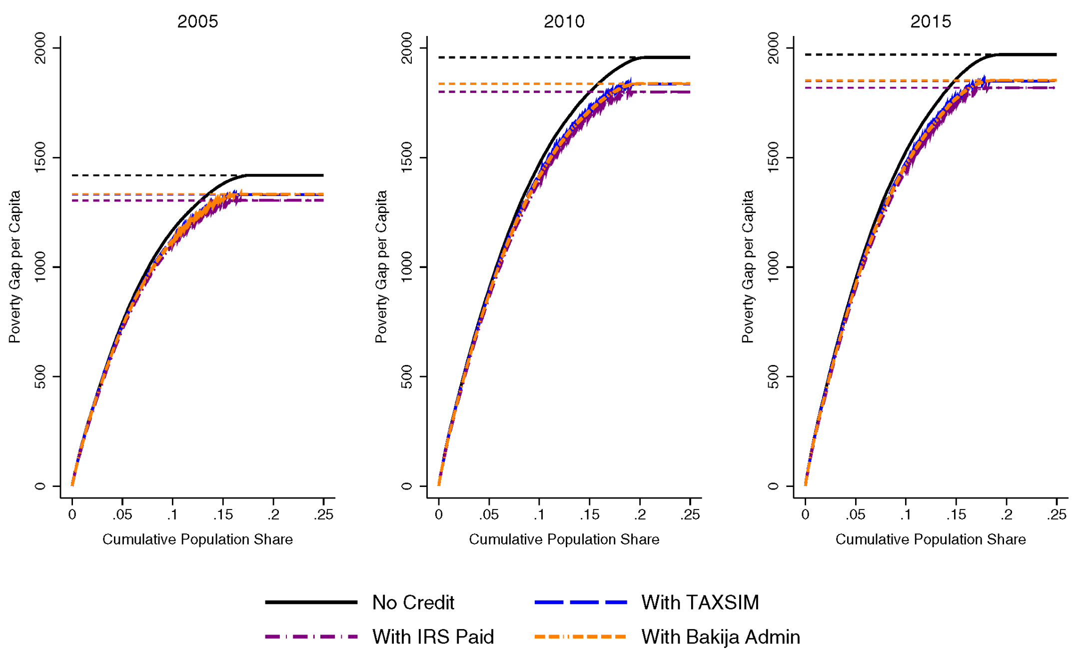


Figure 11. TIP Curve estimates of the number of persons lifted from poverty by EITC. Source: combined EITC/CP0927 recipient files, CPS ASEC, Form 1040, and Form W-2, tax years 2005-2016. Release authorization number CBDRB-FY20-224.

Table 1. Earned Income Tax Credit Parameters, 2005 and 2016 Tax Years

Qualifying Children	Credit rate (percent)	Minimum income for maximum credit	Maximum credit	Phaseout rate (percent)	Phaseout range: Beginning income	Phaseout range: Ending income
2005						
No children	7.65	5,220	399	7.65	6,530	11,750
One child	34	7,830	2,662	15.98	14,370	31,030
Two children	40	11,000	4,400	21.06	14,370	35,263
2016						
No children	7.65	6,610	506	7.65	8,270	14,880
One child	34	9,920	3,373	15.98	18,190	39,296
Two children	40	13,930	5,572	21.06	18,190	44,648
Three children	45	13,930	6,269	21.06	18,190	47,955

Source: Brookings/Urban Tax Policy Center http://www.taxpolicycenter.org/statistics/eitc-parameters

Table 2. Summary of EITC estimates and relevant sample

Estimate	Inputs	Sample
CPS	Marital status, family relationships, and characteristics of head to identify potential filing units; survey reported earnings and income; imputations from Statistics of Income (SOI) for capital gains and itemized deductions.	All CPS ASEC family heads
TAXSIM	Input file of 21 variables on marital status, age, number of dependents, and other family head characteristics; survey reported earnings and income; imputations from SOI on capital gains and itemized deductions. Uses FORTRAN and an FTP protocol to process the input file.	All CPS ASEC family heads
Bakija	Input file of 47 variables on marital status, age, number of dependents, and other characteristics; survey reported earnings and income. Uses SAS programs and ASCII matrices to process the input file.	All CPS ASEC family heads
Bakija Admin	Same as *Bakija*, except using administrative records values of income and earnings in place of survey values when available.	All CPS ASEC family heads
IRS Eligible	Marital status, family relationships, characteristics of head, and survey-reported earnings and income. Each survey value is iteratively replaced with an administrative record (from 1040, W-2, or the EITC recipient file), when available for earnings, income, number of dependents, marital status, and other values based on a set of business rules.	CPS ASEC family heads who receive a unique identifier and who do not have imputed earnings or income.
IRS Paid	The EITC recipient file. A CPS ASEC family head must receive a unique identifier (about 90 percent of CPS ASEC heads) and appear in the recipient file to be considered *IRS Paid.*	All CPS ASEC family heads who receive a unique identifier
IRS Paid and Eligible	Same as *IRS Eligible* plus an appearance in the EITC recipient file.	CPS ASEC family heads who receive a unique identifier and who do not have imputed earnings or income.

Table 3. Ratio of Alternative EITC Model Payments to the Internal Recipient File for Tax Year 2006

	CPS	*TAXSIM*	*Bakija*	*Bakija Admin*	*IRS Paid*	*IRS Eligible*	*IRS Paid & Eligible*
Filing Status							
Head of household							
Total benefits	0.57	0.56	0.56	0.59	0.70	0.63	0.57
Number of recipients	0.60	0.62	0.61	0.65	0.68	0.73	0.62
Joint							
Total benefits	1.05	1.06	1.07	1.10	1.46	1.01	0.83
Number of recipients	1.11	1.16	1.17	1.20	1.42	1.15	0.94
Single							
Total benefits	0.47	0.51	0.46	0.61	1.33	0.22	0.13
Number of recipients	0.73	0.84	0.65	0.96	0.96	0.82	0.47
Qualifying Children							
Zero Children							
Total benefits	2.21	2.33	2.17	2.64	5.73	0.88	0.53
Number of recipients	1.02	1.14	0.93	1.28	1.34	1.08	0.63
One Child							
Total benefits	0.66	0.63	0.63	0.66	0.86	0.69	0.60
Number of recipients	0.65	0.68	0.67	0.71	0.81	0.79	0.67
Two Children							
Total benefits	0.62	0.63	0.63	0.66	0.78	0.68	0.59
Number of recipients	0.70	0.72	0.72	0.76	0.79	0.78	0.66
Total							
Total benefits	0.68	0.68	0.67	0.72	0.95	0.69	0.59
Number of recipients	0.75	0.80	0.85	0.75	0.92	0.85	0.66

Source: combined EITC/CP0927 recipient files, CPS ASEC, Form 1040, and Form W-2, tax year 2006. Counts from the recipient file provide the denominator, with recipient-file characteristics providing the cell definitions; the numerator is the number of persons eligible via the estimation method described in the text, with characteristics defined using the *IRS Eligible* estimation procedure. Release authorization number CBDRB-FY20-271.

Table 4. Ratio of Alternative EITC Model Payments to the Internal Recipient File for Tax Year 2016

	CPS	*TAXSIM*	*Bakija*	*Bakija Admin*	*IRS Paid*	*IRS Eligible*	*IRS Paid & Eligible*
Filing Status							
Head of household							
Total benefits	0.52	0.52	0.52	0.55	0.69	0.62	0.55
Number of recipients	0.55	0.56	0.54	0.60	0.66	0.70	0.60
Joint							
Total benefits	1.11	1.09	1.10	1.06	1.33	1.09	0.86
Number of recipients	1.16	1.14	1.16	1.15	1.29	1.18	0.94
Single							
Total benefits	0.43	0.45	0.41	0.51	0.91	0.19	0.12
Number of recipients	0.64	0.68	0.54	0.84	0.79	0.74	0.47
Qualifying Children							
Zero Children							
Total benefits	2.24	2.23	2.05	2.41	4.53	0.91	0.60
Number of recipients	0.91	0.95	0.79	1.13	1.10	0.99	0.63
One Child							
Total benefits	0.56	0.56	0.55	0.58	0.77	0.65	0.56
Number of recipients	0.58	0.58	0.57	0.62	0.73	0.71	0.61
Two Children							
Total benefits	0.60	0.60	0.60	0.60	0.74	0.66	0.56
Number of recipients	0.68	0.68	0.68	0.69	0.75	0.76	0.64
Total							
Total benefits	0.70	0.71	0.71	0.72	0.79	0.76	0.63
Number of recipients	0.78	0.79	0.79	0.81	0.81	0.89	0.73

Source: combined EITC/CP0927 recipient files, CPS ASEC, Form 1040, and Form W-2, tax year 2016. See caption for Table 3 for further details.

Table 5. Characteristics of CPS ASEC Tax Unit Heads by Eligibility and Recipiency

	Non-eligible non-receiver	Eligible receiver	Eligible non-receiver	Non-eligible receiver	Total
EITC Amount (*IRS Eligible*)	0	1,945	1,082	0	250.00
		(5.236)	(9.442)		(0.980)
EITC Amount (*IRS Paid*)	0	2,270	0	2,016	331.20
		(5.487)		(9.164)	(1.155)
EITC Amount (*Bakija*)	43.44	1,586	806.60	238.70	246.30
	(0.441)	(5.726)	(9.656)	(4.778)	(1.000)
Survey Annual Earnings ('000s)	46.00	20.00	17.00	32.00	42.00
	(0.088)	(0.068)	(0.169)	(0.294)	(0.075)
IRS Annual Earnings ('000s)	44.00	17.00	11.00	27.00	40.00
	(0.329)	(0.037)	(0.074)	(0.195)	(0.269)
Survey Annual Income ('000s)	52.00	21.00	18.00	34.00	47.00
	(0.093)	(0.068)	(0.169)	(0.297)	(0.078)
IRS Annual Income ('000s)	64.00	18.00	12.00	32.00	56.00
	(0.404)	(0.036)	(0.076)	(0.172)	(0.330)
Less than HS	0.114	0.181	0.228	0.192	0.128
	(0.000)	(0.001)	(0.003)	(0.002)	(0.000)
High school grad	0.284	0.372	0.338	0.374	0.299
	(0.001)	(0.002)	(0.003)	(0.003)	(0.001)
Some college	0.275	0.323	0.285	0.298	0.281
	(0.001)	(0.002)	(0.003)	(0.003)	(0.001)
College degree	0.327	0.123	0.150	0.136	0.292
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)
Married	0.406	0.340	0.295	0.400	0.395
	(0.001)	(0.002)	(0.003)	(0.003)	(0.001)
Single	0.594	0.660	0.705	0.600	0.605
	(0.001)	(0.002)	(0.003)	(0.003)	(0.001)
Number of qualifying children	0.454	1.353	0.999	0.692	0.580
	(0.001)	(0.003)	(0.008)	(0.006)	(0.001)
Number of persons	2.238	3.132	2.721	2.625	2.367
	(0.002)	(0.005)	(0.011)	(0.009)	(0.002)
Obs.	763,000				
Weighted N in millions	1,890				

Source: combined EITC/CP0927 recipient files, CPS ASEC, Form 1040, and Form W-2, tax years 2005-2016. Release authorization number CBDRB-FY20-224. Incomes in real $2009 PCE

Online Appendix---Not for Publication

Data Appendix

The survey data used are yearly internal ASEC files from survey years 2006 to 2017, corresponding to tax years 2005-2016. The ASEC is a nationally representative survey of about 90,000 households, conducted as a supplement to the monthly CPS labor force survey in March of each year (with some interviews conducted in February and some in April). The tax data included in the project are, for each year, Form 1040 individual income tax records, the EITC recipient file, the CP09/27 file (a record of taxpayers sent a notice from the IRS about their potential EITC eligibility), and Form W-2 wage and tax statement. The Census Bureau receives tax records from the IRS to calculate and report on the take-up rate of the EITC, with the calculation of the denominator dependent upon survey data that is representative of the U.S. population. The survey data allow us to determine the members of the population who appear to be eligible, regardless of whether they file a Form 1040. The process of eligibility modeling and take-up calculation is reported in detail for tax year 2005 in Plueger (2009). The process, briefly described below, has changed somewhat in subsequent years, mainly in the refinement of income measurement.

The survey data in the ASEC are matched at the individual level for the corresponding tax year with the IRS data. The records are made linkable using a process whereby individuals in each data set were given a unique, protected identification key, or PIK. When a Social Security Number (SSN) is available in a data set (such as all of the IRS records used in this project), the PIK is assigned based on SSN. Identifier placement is close to 100 percent in the case of administrative tax records with an SSN. The ASEC stopped collecting SSNs as of the 2006 survey year, and thus personally identifiable information such as name, address, and date of birth is used in probabilistic matching against a reference file to assign PIKs (Wagner and Layne

2014). Personal information is then removed from each data set before it may be used for research purposes. For the IRS EITC estimation project, we also remove persons whose income and wage values were imputed in the ASEC, as initial EITC eligibility determination, which uses only the survey data, is dependent on these values. We then reweight the data based on the probability that an observation received a PIK and did not have imputed or edited information.

A. Tax Simulators

The first survey-based approach to estimating the EITC is the method long used within Census (*CPS*). This measure is part of a larger package of simulated tax variables that Census has provided to users of the internal- and public-use ASEC, first developed in the 1980s and revised in survey year 2004 (U.S. Bureau of the Census 1993; O'Hara 2004). The CPS model first computes payroll tax liability for each person with earned income. It then constructs potential filing units (single, joint, head of household, dependent filer) based on marital status and household relationships (i.e., a household may have multiple tax filing units from related or unrelated subfamilies/individuals). These units are then statistically matched to the Statistics of Income (SOI) Public Use File to impute capital gains and itemized deductions, which are not collected in the ASEC. This provides input to compute initial federal adjusted gross income (AGI) and tax liability and credits, which in turn are used to construct state tax payments and credits, and then final federal taxes are computed using the estimated state tax payments as a deduction (for simulated itemizers). For our purposes, we focus on the EITC variable (eit_cred).

The second modeling strategy, widely used by many researchers, employs the NBER's TAXSIM model (*TAXSIM*). This tax simulator uses up to 21 input variables derived from source data that reflect tax unit characteristics, including marital status, age of the primary taxpayer (and secondary if present), along with their wage and salary income, state of residence, number and ages of dependents (for calculation of EITC, child tax credit, etc.), and other taxable and nontaxable income, and potentially deductible expenses (home mortgage interest, property tax).[19, 20] Similar to the CPS model, TAXSIM uses the SOI Public Use File to impute itemized deductions for each filing unit and compares these to the standard deduction for each filing unit to determine whether the taxpayer is assumed to itemize, along with an iterative procedure between federal and state tax payments. There are a possible 38 outputs provided by TAXSIM, with the federal EITC (v25) the output of focal interest to our project.

The third simulator we examine is the tax model developed by Jon Bakija (*Bakija*). His simulator differs from TAXSIM in its programming language (SAS versus FORTRAN) and in the larger number of variables that must be inputted, but it otherwise is similar. Comparisons between TAXSIM and the Bakija models indicate that, conditional on receiving comparable input data, their tax output is highly correlated (>0.9 for most output variables considered in our analysis).

For *TAXSIM* and *Bakija*, we use the ASEC to create tax units and assign dependents to filers. Rather than use Census-constructed input variables such as filing status, we construct our own.[21] The first task is to assign the heads and spouses (if applicable) for each potential filing unit identified by a unique ID based on household sequence number, family sequence number, family position, and family type. The head can be of the primary family, of a related subfamily,

[19] We use version 9 of TAXSIM that has been installed for research purposes with the internal ASEC at the Census Bureau. It is the same as Internet TAXSIM (v9) available at http://users.nber.org/~taxsim/taxsim-calc9/. There is also available to public users a new version, TAXSIM 27, that inputs 27 variables instead of the 21 used in TAXSIM 9. A copy of the code to prep the ASEC can be found at https://sites.google.com/site/jamesziliak/Home/Research

[20] TAXSIM is traditionally used with survey data such as the ASEC, ACS, SIPP, PSID, and others, but is also employed to flesh out information from administrative records when full tax information is missing.

[21] A Stata DO file with the code that prepares the ASEC data for input into TAXSIM is available from the authors at https://sites.google.com/site/jamesziliak/Home/Research/. Our approach is an update of that made available by Judith Scott-Clayton to researchers using the ASEC as inputs to TAXSIM via the Stata interface. See http://users.nber.org/~taxsim/to-taxsim/cps/cps-clayton/.

of an unrelated subfamily, or a primary individual. We also allow for dependent filers. We then construct a variable for the number of dependents based on age of the child and relationship to head, including those between ages 18 and 24 who are full-time students (and thus can be claimed as dependents for the EITC) as well as foster children. ASEC observations are assigned as nonfilers if they are a dependent child, as single if they are unmarried and have no dependents, as head of household if they are unmarried and with dependents, and as joint filers if they are married with or without children. Wage and salary income is constructed from ASEC variables, including farm and self-employment earnings, while the other taxable and nontaxable income sources are assigned to the taxpayer according to input requirements of TAXSIM or Bakija. Each primary and secondary taxpayer is run through the simulator, but the tax values of the primary taxpayer are the only ones retained to avoid double counting.

B. Administrative Tax Models

We compare these simulators to three others that use survey data on the filing unit in combination with actual IRS tax records. The first, *IRS Eligible*, consists of tax units who are estimated to be IRS eligible based on information in the ASEC supplemented by administrative records, regardless of whether they file a return or not. That is, we assume 100 percent take-up, using administrative IRS income amounts from the Form 1040 and W-2 for those who file and get paid, and a predicted amount based on program parameters, survey information, and W-2 information for those who do not file. We follow the method outlined in Plueger (2009), where the assignment of persons to tax units and the identification of filers and qualifying children is essentially the same as the method used for *TAXSIM* and *Bakija*. However, throughout the modeling process, eligibility is refined by substituting values from the matched 1040 data whenever available. We presume the reference person in each family to be the primary tax filer for the identified tax unit. In a later adjustment, the tax information on eligibility is transferred from the householder to the spouse if it was the spouse who filed. Variables on tax unit earnings, income, and dependent support requirements (which, in the case of the EITC, is based on where a child lived for the tax year) are first taken from the survey data. Then, values from the tax data, when available, are swapped in for the survey values, and eligibility is refined based on these new values.[22]

The second ASEC-IRS model consists of those tax units in actual receipt of administrative EITC payment by the IRS regardless of eligibility, which we denote as (*IRS Paid)*. The *IRS Paid* model is informative as it provides estimates of the total dollar amounts of EITC credits circulating in the economy, regardless of whether those payments are correct. Finally, we are interested in EITC receipt for ASEC respondents who are jointly IRS eligible and paid the EITC (*IRS Paid & Eligible*). A comparison of the *IRS Paid* and *IRS Paid & Eligible* sample provides an estimate of the amount of incorrect EITC payments made in a given year, while the ratio of the latter to the former provides an estimate of EITC take-up.

Table A1 provides summary statistics on the sample, the mean eligibility rates, and the mean values of the estimators. Each of the survey-based estimators have similar mean eligibility values, while the administrative tax models show lower eligibility rates and dollar values.

Table A1. Summary Statistics for Linked ASEC-IRS Data, Harmonized Sample of Potential Tax Filers

Full sample	Mean	Standard Err.	95% Confidence Interval	
Reweighted (Obs.= 764 thou., N=1,863 mil.)				
EITC Receiver (*CPS*)	0.12	0.00	0.13	0.12
EITC Receiver (*TAXSIM*)	0.13	0.00	0.13	0.13
EITC Receiver (*Bakija*)	0.12	0.00	0.13	0.12
EITC Receiver (*IRS Eligible*)	0.14	0.00	0.14	0.14

[22] Once weighted, filers identified in the ASEC represent 94 percent of 1040 filers and 93 percent of W-2 observations. It should be noted that determinations of eligibility are based on assessments of the household roster as reported in the ASEC, and not on audits. When we say tax units are paid regardless of eligibility, we do not imply that these tax units would be found noncompliant or fraudulent if subjected to an audit.

EITC Receiver (*IRS Paid)*	0.15	0.00	0.15	0.15
EITC Amount (*CPS*)	244.50	1.47	247.40	241.60
EITC Amount (*TAXSIM*)	247.80	1.47	250.70	244.90
EITC Amount (*Bakija*)	246.30	1.46	249.20	243.50
EITC Amount (*Bakija Admin*)	256.40	1.35	259.00	253.70
EITC Amount (*IRS Eligible*)	250.00	1.41	252.70	247.20
EITC Amount (*IRS Paid)*	331.20	1.78	334.60	327.70
EITC Amount (*IRS Paid & Eligible)*	243.30	1.39	246.10	240.60
Annual Income	56250.00	464.50	57160.00	55340.00
Annual Earnings	39680.00	349.50	40370.00	39000.00
Annual Federal tax	6037.00	37.77	6111.00	5963.00
Less than HS	0.13	0.00	0.13	0.13
High school grad	0.30	0.00	0.30	0.30
Some college	0.28	0.00	0.28	0.28
College degree	0.29	0.00	0.29	0.29
Married	0.39	0.00	0.40	0.39
Single	0.61	0.00	0.61	0.60
White	0.79	0.00	0.79	0.79
Black	0.14	0.00	0.14	0.13
American Indian or Alaska Native	0.01	0.00	0.01	0.01
Asian	0.05	0.00	0.05	0.05
Other	0.02	0.00	0.02	0.02
Hispanic	0.15	0.00	0.15	0.15
Sex (1=male)	0.60	0.00	0.61	0.60
Age	47.98	0.02	48.01	47.94
Number of children <18	0.51	0.00	0.51	0.51

Family means				
EITC Amount (*CPS*)	235.30	1.48	238.20	232.40
EITC Amount (*TAXSIM*)	240.10	1.48	243.00	237.20
EITC Amount (*Bakija*)	240.00	1.48	242.90	237.10
EITC Amount (*Bakija Admin*)	1019.00	6.85	1033.00	1006.00
EITC Amount (*IRS Eligible*)	298.80	1.77	302.30	295.40
EITC Amount (*IRS Paid*)	407.40	2.39	412.10	402.80
EITC Amount (*IRS Paid & Eligible*)	290.90	1.75	294.30	287.50
Number of persons	2.37	0.00	2.38	2.36
Number of workers	1.25	0.00	1.25	1.24
Number of qualifying children	0.58	0.00	0.58	0.58
Total income ($000s)	63980.00	197.40	64370.00	63600.00

Source: combined EITC/CP0927 recipient files, CPS ASEC, Form 1040, and Form W-2, tax years 2005-2016. Numbers have been rounded to comply with the Census Bureau's disclosure avoidance guidelines. Dollar amounts are in real 2009 dollars using the personal consumption expenditure deflator. Release authorization number CBDRB-FY20-224.

C. Comparison of internal EITC file to SOI reports

Our comparison of estimates begins with providing an assessment of weighted CPS persons versus the EITC recipient file; here, we provide evidence that the EITC recipient file submitted to Census aligns closely to published aggregates from the IRS's Statistics of Income, and thus can serve as the benchmark standard to measure the alternative modeling strategies. Table A2 contains both the numbers of recipients (in millions) and the benefits paid (in millions of nominal dollars) for tax year 2006 as reported in SOI public reports, as well as comparable numbers from the internal EITC recipient file submitted by IRS to Census annually for estimation of EITC eligibility and take-up. As in Meyer (2010) we present the statistics broken down by number of qualifying children, and also use tax year 2006 as in his study. The internal file covers 97 percent of EITC dollars and 100 percent of recipients reported publicly. There are some differences by number of children, with the internal file finding a slightly higher proportion of EITC recipients with zero children. This slight discrepancy may be due to amended filings, which may be reflected differently in the two sources depending on the vintage of the source. Overall, however, the internal recipient file is representative of the full population of EITC claims.

Table A2. Comparison of Internal Recipient File to Aggregates in Public-Use IRS Statistics of Income, Tax Year 2006
(numbers in millions)

	(a) Public Use Aggregates		(b) Internal Recipient File		(c) Ratio
Zero Children					
Total benefits	$1,142	2.57%	$1,133	2.63%	0.99
Number of recipients	4.81	20.88%	5.11	22.17%	1.06
One Child					
Total benefits	$16,078	36.22%	$15,750	36.57%	0.98
Number of recipients	8.75	37.98%	8.66	37.57%	0.99
Two Children					
Total benefits	$27,168	61.21%	$26,180	60.80%	0.96

Number of recipients	9.49	41.19%	9.28	40.26%	0.98
Total					
Benefits	$44,388	100%	$43,070	100%	0.97
Recipients	23.05	100%	23.05	100%	1.00

Source: EITC/CP0927 recipient files, 2006, and SOI public reports (U.S. Department of the Treasury, Individual Income Tax Returns 2004). Numbers in column (b) have been rounded to comply with the Census Bureau's disclosure avoidance guidelines.

D. Comparison of Official Poverty Estimates to Alternative Estimates Inclusive of EITC

In Table A3 we present point estimates of the official poverty rate, along with estimates of poverty after inclusion of the EITC. In panel A we use the full ASEC, inclusive of persons with imputed earnings and incomes and those with and without a PIK. The column labeled ASEC replicates the official poverty rate found in annual P-60 reports. Panel B restricts the analysis to those who respond to the earnings and income questions, and with a PIK to the tax data. The estimates of the number of persons lifted out of poverty by the EITC in the figures found in the text of the paper are computed by subtracting the population-weighted number of persons in each post-EITC estimate less the population-weighted estimates using the official poverty measure.

Table A3. Effect of EITC on the Official Poverty Rate by Alternative EITC Eligibility Models and Samples

A. Full Sample and ASEC Weight					
Year	ASEC	*CPS*	*TAXSIM*	*Bakija*	*IRS Paid*
2005	12.6	11.3	11.3	11.2	11.7
2006	12.3	11.0	11.0	10.9	11.4
2007	12.5	11.1	11.0	11.0	11.5
2008	13.2	12.0	11.9	11.9	12.4
2009	14.3	12.6	12.6	12.6	13.2
2010	15.1	13.4	13.4	13.4	14.0
2011	15.0	13.2	13.2	13.2	13.8
2012	15.0	13.3	13.3	13.2	13.8
2013	14.7	13.0	12.9	12.9	13.6
2014	14.8	13.1	13.0	13.0	13.6
2015	13.5	12.1	12.1	12.1	12.6

2016	12.7	11.3	11.3	11.3		11.8		

B. Harmonized Sample and Reweighted ASEC Weight for Nonresponse and Link

Year	ASEC	*CPS*	*TAXSIM*	*Bakija*	*Bakija Admin*	*IRS Paid*	*IRS Eligible*	*IRS Paid & Eligible*
2005	12.2	11.3	11.3	11.3	11.4	11.3	11.4	11.4
2006	12.2	11.4	11.4	11.4	11.5	11.3	11.4	11.4
2007	12.1	11.1	11.1	11.1	11.3	11.1	11.2	11.3
2008	12.7	11.8	11.8	11.8	11.9	11.7	11.9	11.9
2009	14.2	12.9	12.9	12.9	13.0	12.9	13.0	13.1
2010	15.2	14.0	14.0	14.0	14.2	14.0	14.1	14.1
2011	15.2	13.9	13.9	13.9	14.0	13.9	13.9	13.9
2012	15.1	13.9	13.9	13.9	14.0	13.8	13.9	13.9
2013	15.0	13.9	13.8	13.8	14.0	13.8	13.9	13.9
2014	15.2	14.0	14.0	14.0	14.2	14.0	14.0	14.1
2015	14.0	13.0	13.0	13.0	13.1	13.0	13.0	13.0
2016	12.6	11.7	11.7	11.7	11.9	11.7	11.7	11.7

Source: combined EITC/CP0927 recipient files, CPS ASEC, Form 1040, and Form W-2, tax years 2005-2016. Release authorization number CBDRB-FY20-224.

E. Estimation of the population of likely ineligibles

The EITC estimation project at the Census Bureau is a joint collaboration with the IRS. Briefly, the goal for IRS has been the estimation of EITC take up versus EITC eligibility. The agency recognized that this estimate requires data that capture household structure, earnings, and income for the full population (when weighted), regardless of filing status. The broad incentive for Census was the availability of what it considered confirmatory data for calibrating its internal tax model and comparing reports of earnings and income to “true” reports. There is some controversy over treating 1040 information as more “truthful” than survey responses; however, for many lower-income tax filers, the parameters that go into EITC eligibility receipt are reported by third parties, making these elements more reliable.[23]

[23] Examples are W2 wage reports and 1099 reports of investment/dividend income.

For household structure, both IRS and Census assume the responses from the ASEC, versus dependent claiming in the 1040, as more likely to reflect the truth. A criticism of the estimation strategy questions that this assumption is valid. The issue of concern is that the ASEC's questions are asked in March, and household structure may have changed since the preceding tax year. IRS and Census argue that the complicated nature of credit and dependent qualifications for 1040 filing, leading to the accidental claiming of children outside of the household for EITC, makes the 1040 report more questionable. A child who is claimed for any credit or exemption on a 1040 cannot be claimed on another's 1040 for a credit or exemption, despite the fact that the rules are different (support versus residency) for each credit. These varying rules lead to confusion among tax filers, who may assume that any child they support—regardless of residency—also qualify them for EITC. In considering which circumstance was more likely (household change versus accidental claiming of non-resident children), IRS and Census researchers determined the latter was the more problematic issue (see Liebman, 2000; McCubbin, 2000; and Holtzblatt and McCubbin, 2003, for the issue of misreporting children for EITC; each paper includes an analysis of IRS audit data).

To further assess the question of households changing structure between the tax year and March, we exploit the longitudinal nature of the ASEC, where approximately half the sample can be linked to their preceding March responses (see Madrian & Lefgren, 1999, and Feng, 2008 for more on linking March samples). Because of the unique identifier we place on the data, we can perform this match either using the PIK or using the probability match as one would with the public-use ASEC. Table A4 reports how well personal attributes correspond based on the type of match. The first column reports the attributes for those persons that have a successful match using both the PIK and probability match (88.1% of the sample); the second column is for those

Table A4: Two-year ASECs, by PIK match or probability match

Linked years		Type of match (1) Both	(2) PIK only	(3) Prob only
2005-2006	Race	1.000	0.967	0.999
	Hispanic	0.998	0.987	0.995
	Sex	1.000	0.843	1.000
	percent	*89.97*	*1.01*	*9.02*
2006-2007	Race	1.000	0.949	1.000
	Hispanic	0.998	0.988	0.996
	Sex	1.000	0.887	1.000
	percent	*89.63*	*0.76*	*9.61*
2007-2008	Race	1.000	0.976	1.000
	Hispanic	0.999	0.985	0.997
	Sex	1.000	0.836	1.000
	percent	*88.2*	*0.73*	*11.07*
2008-2009	Race	1.000	0.896	1.000
	Hispanic	0.999	0.984	0.994
	Sex	1.000	0.828	1.000
	percent	*88.3*	*0.6*	*11.1*
2009-2010	Race	1.000	0.961	0.999
	Hispanic	0.999	0.984	0.995
	Sex	1.000	0.832	1.000
	percent	*89.35*	*0.61*	*10.04*
2010-2011	Race	1.000	0.933	1.000
	Hispanic	0.999	0.986	0.993
	Sex	1.000	0.870	1.000
	percent	*90.22*	*0.66*	*9.13*
2011-2012	Race	1.000	0.936	0.999
	Hispanic	0.999	0.976	0.995
	Sex	1.000	0.889	1.000
	percent	*89.52*	*0.67*	*9.81*
2012-2013	Race	1.000	0.947	0.997
	Hispanic	0.999	0.971	0.996
	Sex	1.000	0.840	1.000
	percent	*88.6*	*0.71*	*10.68*
2013-2014	Race	1.000	0.992	0.999
	Hispanic	0.999	0.996	0.995
	Sex	1.000	0.991	1.000
	percent	*78.17*	*10.97*	*10.86*
2014-2015	Race	1.00	0.968	1.00
	Hispanic	0.999	0.973	0.997
	Sex	1.00	0.855	1.00

	percent	*87.76*	*0.8*	*11.44*
2015-2016	Race	1.00	0.978	1.00
	Hispanic	0.998	0.992	0.995
	Sex	1.00	0.837	1.00
	percent	*87.33*	*0.69*	*11.98*
N		588,000	10,000	65,500
		88.10	1.51	10.39

Source: combined EITC/CP0927 recipient files, CPS ASEC, Form 1040, and Form W-2, tax years 2005-2016. Counts are rounded to conform to the U.S. Census Bureau's disclosure avoidance practices. Release authorization number CBDRB-FY20-271.

for whom a PIK match is made, but the probability match is not successful (1.5%); and the third column is for those with a successful probability match but not a successful PIK match (10.4%). These latter cases likely reflect the structure of the CPS, where households form the sample but the families within a household may change from March to March. Because gender is used as a matching variable, it is not surprising that gender is consistent for a matched individual 100 percent of the time for both column 1 and column 3. We also use broad race categories for matching; there is close to a 100 percent chance in all years that exact race also matches between the two years. Hispanic correspondence occurs for 98 to 99 percent. This might be due to a few persons reporting differently in the two different waves, as Hispanic identity seems to be a more fluid characteristic than race and sex (Liebler et al., 2017). In contrast, when looking at PIK matches that don't match based on probability matching, rates of correspondence range from 99 percent (for race and Hispanic origin) and 83 percent (for sex). The total number of matches constitute about 30 percent of the base-year number of observations.

Focusing on the apparent error in child claiming, about 1,940 filers claim more children for EITC than appear in their household. Of these, the vast majority (70 percent) appear to have zero children in the first year's March survey (Table A5).[24] Approximately 400 persons appeared

[24] Many of these also appear to have zero children in the second year's March survey, but for simplicity the table pools claiming disparities according to the number of children claimed in the second year. For example, those who

to have as many or more children in their household in the first year compared with their claiming in the second year (about 21 percent of these filers). The remaining 9 percent have fewer, but non-zero, children in the preceding survey compared with their filing in the second year.

Table A5. Those with survey-determined children < claimed children, by survey-determined children in year 1

Children claimed	Children in survey, year 1			
Year 2	Zero	One	Two	Three
One	900	250	30	20
	75.00	*20.83*	*2.50*	*1.67*
Two	450	150	100	<15
	63.38	*21.13*	*14.08*	*(D)*
Three or more	100	20	30	20
	58.82	*11.76*	*17.65*	11.76
Total	1,400	350	150	(D)
Percent with agreement or +	21%			

Thus the number that are potentially misclassified as incorrectly paid—approximately 400 filers—constitutes about 5.5 percent of all filers we identify as incorrectly paid EITC in the matched sample (thus about 2 percent of all incorrectly paid filers identified in the pooled CPS ASECs). Of course, this is a lower bound for all filers, since we, in theory, can only match one-half of second-year observations to their first-year characteristics and, in realty, only match one-third. On the other hand, we have no way in which to identify changes that occur between March of year one and December of year one—household structure is captured early enough in the tax year that it may not reflect the structure over the full year. An EITC-qualifying child must be resident in the household for more than half the tax year. In short, especially for the purposes of the current analysis, it is unlikely that misclassifying such a small number of likely paid

claim two children in the second year (e.g., tax year 2006) may appear to have zero or one child in their household in the 2007 ASEC to fall under the category "survey-determined children < claimed children."

ineligibles has much of an impact on our estimates. This is especially true for our anti-poverty analysis, since many paid ineligibles share most characteristics with the eligible population, including low income.

References

Feng, S. 2008. “Longitudinal Matching of Recent Current Population Surveys: Methods, Non-Matches and Mismatches.” Journal of Economic and Social Measurement, 33(4): 241-252.

Holtzblatt, J. and McCubbin, J., 2003. “Whose child is it anyway? Simplifying the definition of a child.” National Tax Journal, 56(3): 701-718.

Liebler, C.A., S.R. Porter, L.E. Fernandez, et al. 2017. “America’s Churning Races: Race and Ethnicity Response Changes Between Census 2000 and the 2010 Census.” Demography, 54(1): 259–284.

Liebman, J. B., 2000. Who Are the Ineligible EITC Recipients. National Tax Journal, 53(4), pp. 1165-1185.

Madrian, B., and L. Lefgren. 1999. “A Note on Longitudinally Matching Current Population Survey (CPS) Respondents.” NBER Working Paper 247.

McCubbin, J., 2000. EITC Noncompliance: The Determinants of the Misreporting of Children. *National Tax Journal,* 53(4, Part 2), pp. 1135-1164

Meyer, Bruce D. 2010. “The Effects of the Earned Income Tax Credit and Recent Reforms,” In *Tax Policy and the Economy*, 24(1): 153-180.