THE ANTIPOVERTY IMPACT OF THE EITC: NEW ESTIMATES FROM SURVEY AND ADMINISTRATIVE TAX RECORDS

Maggie R. Jones and James P. Ziliak

We reassess the antipoverty effects of the earned income tax credit (EITC) using unique data linking the Current Population Survey (CPS) Annual Social and Economic Supplement to Internal Revenue Service (IRS) data for the same individuals spanning tax years 2005–2016. We compare EITC benefits from standard simulators to administrative EITC payments and find that the antipoverty estimates of the EITC are countercyclical in terms of number of recipients, with roughly four million people of all ages and 1.9 million children lifted from after-tax poverty in a typical year. We outline how researchers using public data can address discrepancies between survey estimates of the EITC and administrative tax records.

Keywords: tax policy, poverty, administrative data linkage, automatic stabilizer

JEL Codes: H2, I3

I. INTRODUCTION

Among the means-tested transfer programs in the US social safety net, the earned income tax credit (EITC) stands out as one of the largest in terms of expenditure and reach, with spending exceeding $63 billion on more than 25 million individuals and families in 2019 (Moffitt and Ziliak, 2019).[1] Research has shown that the program has led to greater employment of women, improved child achievement, provided more stability of household financial balance sheets, and reduced racial inequality,

Maggie R. Jones: Center for Economic Studies, US Census Bureau, Suitland, MD, USA (margaret.r.jones@census.gov); James P. Ziliak: Department of Economics, University of Kentucky, Lexington, KY, USA (jziliak@uky.edu)

[1] https://www.eitc.irs.gov/partner-toolkit/basic-marketing-communication-materials/eitc-fast-facts/eitc-fast-facts.

Electronically published August 1, 2022

National Tax Journal, volume 75, number 3, September 2022.
 https://doi.org/10.1086/720614

among other outcomes.[2] One of the most highly touted benefits of the EITC is that it has lifted more people out of poverty than any other safety net program for children and nonelderly working households (Ziliak, 2015a; Hoynes and Patel, 2018). Given its record of success in combating poverty, there have been calls from the policy and research communities to expand the program in both eligibility and generosity (Marr, Horton, and Duke, 2017; Duncan and Le Menestrel, 2019; Hoynes, 2019).

The research evidence to date on the antipoverty effects of the EITC underpinning these policy proposals is based upon household survey data. However, a potential challenge is that major household surveys do not collect information on credit eligibility, receipt, or amount. In general, EITC eligibility and dollar amounts are simulated based on survey reports of age, family structure, earnings, income, and limited other information salient to tax liability, much of which may be reported with error. Indeed, survey estimates of aggregate EITC recipients and dollars received fall short of Internal Revenue Service (IRS) reports by about a third (Meyer, 2010), which raises questions about the accuracy of antipoverty estimates derived from survey values. Potentially compounding this error is the maintained assumption in tax simulators of 100 percent take-up of the EITC conditional on eligibility — estimates using administrative (IRS) data place actual take-up rates closer to 80 percent (Scholz, 1994; Plueger, 2009; Jones, 2014). While a 100 percent take-up assumption may tend to overstate antipoverty effects, evidence also exists that the IRS pays some EITC claims that ex post are deemed ineligible (Marcuss et al., 2014), and thus survey simulations may understate the actual impact of the credit because they miss ineligible claims. Ultimately, these errors will function together to over- or understate the EITC's impact on poverty depending on how each type of error interacts with families whose income places them close to or far away from the poverty line.

Our goal in this paper is to provide new estimates of the antipoverty impact of the EITC, and in the process to reconcile survey-based estimates of EITC recipients and dollars with publicly reported IRS aggregates of actual recipients and dollars distributed to taxpayers. This reconciliation is important to provide improved guidance on evidence-based policymaking regarding the EITC, most of which is based from publicly available survey data. We provide insights into how the use of survey data alone might mismeasure the antipoverty impact of the EITC, as well as the challenges of incorporating administrative records into survey-based estimation. Using these insights, we outline strategies for better poverty estimation for researchers who do not have access to IRS records.

We use a unique, internal Census data set linking survey information from the Annual Social and Economic Supplement (ASEC) of the Current Population Survey (CPS) to administrative IRS tax data from Form 1040, W-2 wage statements, and

[2] See, e.g., Meyer and Rosenbaum (2001), Grogger (2003), Dahl and Lochner (2011), Bastian and Michelmore (2018), Jones and Michelmore (2018), Bastian and Jones (2021), Schanzenbach and Strain (2021), and Hardy, Hokayem, and Ziliak (2022). Hotz and Scholz (2003) and Nichols and Rothstein (2016) provide comprehensive surveys of research on the EITC.

the EITC recipient file to assess the effects of simulated EITC versus actual credit values on after-tax and transfer poverty rates for tax years 2005–2016. Against the benchmark of actual EITC receipt, we focus on two EITC simulators widely used by the research and policy communities and available in the public domain: one produced by the National Bureau of Economic Research's TAXSIM model, and the second by the Census Bureau as part of its annual release of the ASEC (CPS model).

We begin our analysis by estimating how many people (both all ages and children) each simulated EITC payment lifts out of poverty. Our resource measure is an after-tax and in-kind transfer income measure akin to that recommended by the National Academy of Sciences (Citro and Michael, 1995) and implemented by the Census Bureau in the supplemental poverty measure (SPM). In the full ASEC sample with the tax simulators, we find that the number lifted out of poverty by the EITC is countercyclical from a headcount perspective — increasing from just more than four million people in the mid-2000s, then rising to about 5.5 million during the years surrounding the Great Recession, only to fall back to just more than four million by 2016, coinciding with the economic recovery. However, we find that in a typical year the tax simulators yield estimates of people lifted from poverty by the EITC that are higher by at least 35 percent compared with the actual credit paid, and by more than 45 percent among children.

The next part of our analysis focuses on identifying the sources of discrepancy between the administrative and tax simulator estimates of the EITC on poverty, including what roles missing linkages to tax records and missing earnings reports have on the antipoverty estimates of the EITC. When we restrict the sample to those who are linkable to the IRS data and who provide complete survey reports on earnings, and then reweight the resulting sample with estimated inverse probability weights (IPW) in a bid to retain population representativeness, we find that the discrepancies between linked administrative and tax simulator antipoverty estimates of the EITC are largely eliminated, differing by 5–8 percent for all individuals, and by only 1–3 percent in the years after the Great Recession. Our estimates are that in an average year, the EITC lifts about four million people — including 1.9 million children — out of poverty.

Using the linked sample of survey earnings respondents, we then decompose the full distribution of EITC payments from survey and administrative inputs, focusing on the number of qualifying children, filing status, and earnings. We also explore the role of self-employment income on the EITC distribution. Here we find that significantly more actual EITC payments flow to childless tax units than predicted by tax simulators. However, those payments still flow to low-income tax units and thus appear target efficient at both the taxpayer and household levels. The discrepancy between the distributions of actual EITC payments and those from tax simulators is accounted for by differences in administrative and survey reports on earnings and qualifying children — including self-employment income — and not tax-filing status. However, we closely approximate the number of people lifted from poverty by the actual EITC using the tax simulators regardless of survey reports of earnings and children, or with these substitutions from administrative records, once we focus on the linked sample of respondents. The assumption of 100 percent take-up in the simulators seems to balance out the possibly ineligible actual EITC payments, yielding comparable antipoverty effects.

In the next section, we provide a brief overview of the EITC and research on estimating the antipoverty effects of the credit from survey-based tax simulators. Section III then describes the data used and how we construct our measure of after-tax and in-kind transfer household income. Section IV discusses the results, first presenting estimates of the antipoverty effects of the EITC, and then followed by a detailed examination of the distribution of EITC payments by number of qualifying children, tax-filing status, and replacing survey input values with their corresponding records from tax data. The final section concludes with recommendations for research on the EITC when links to tax data are not available.

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 substantially increased the income eligibility for taxpayers 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 tax units grew fourfold to 27.4 million, with the average inflation-adjusted credit growing tenfold to $2,437.[3]

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

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

Qualifying Child	Credit Rate (%)	Minimum Income for Maximum Credit	Maximum Credit	Phase-Out Rate (%)	Phase-Out Range Beginning Income	Phase-Out 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.

The growth in the credit has generated a voluminous research literature, including estimates of the antipoverty impact. 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).[4] Ziliak (2015a), using public-release versions of the ASEC and the CPS tax model, estimated that the EITC lifted about four million people out of poverty (based on the official poverty measure) per year in the decade prior to the Great Recession in 2008 and more than five million at the peak of the recession.[5] The Center on Budget and Policy Priorities (2016) estimated that 6.5 million people were lifted out of SPM poverty in 2015 using the public ASEC with the CPS tax model.[6] Hoynes and Patel (2018), using the public ASEC along with TAXSIM, estimated that the EITC lifted 3.4 million children in single-mother families out of poverty in 2012, including both the mechanical and the behavioral effects of the credit. Hoynes and Patel (2018) used the official poverty line with a broader

[4] A recent paper by Kleven (2019) suggested no extensive-margin labor supply response to the EITC, but Schanzenbach and Strain (2021) challenge the negative result in Kleven, finding a sizable extensive-margin response.

[5] The official poverty measure is a pretax measure that includes cash transfers but not in-kind transfers or tax payments and credits.

[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. We report a similar result below.

resource definition that included some in-kind transfers and tax payments and credits. We use an after-tax and transfer resource definition similar to that of Hoynes and Patel (2018), but we use the more comprehensive SPM thresholds for the poverty line. As our focus is on credit measurement, we only examine the mechanical effect of the credit.

A challenge facing research on the EITC is that none of the major household surveys in the United States 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 (from an employer or from self-employment), adjusted gross income, interest income, age of filer, age and number of qualifying children, and tax-filing status. Since the 1980s, the Census Bureau uses available survey information in the ASEC to construct tax units and simulate federal, state, and payroll tax liability, including the EITC. They make the output available in public-release versions of the ASEC but suppress many of the input variables such as who is estimated to be in the tax unit. Each user of the TAXSIM model therefore has to independently compile this information from the survey. This is both a strength and weakness of TAXSIM; a strength because it can be incorporated across multiple data sets and platforms, and a weakness in that the user must make complex decisions based on the available data to construct credible tax estimates. While most surveys have enough information to create pointers to family and household relationships, these may not translate directly into tax-filing units; for example, a family or household may contain multiple filing units. In addition, determining who does and who does not meet the EITC "qualifying child" test is difficult. Children under age 19 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 POVERTY MEASURES

The data we use derive from a joint statistical agreement between the Census Bureau and the IRS. The survey data are yearly internal-to-Census ASEC files from 2006 to 2017, linked at the individual level for the corresponding tax year with the IRS data (i.e., the 2005–2016 tax years).[7] The ASEC is a supplement to the monthly CPS survey conducted in March of each year that contains detailed information on earnings and incomes from the prior year, employment, family structure, among other socioeconomic outcomes. Because the ASEC is a stratified random sample, the Census provides weights to make the sample representative of the US population. The tax data include 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.

A. Sample Selection

We define the poverty population similar to that used in the Census's SPM (Renwick and Fox, 2016; Fox, 2019), which is broader than the population used in official poverty statistics. The official poverty population is based on family membership emanating from marriage, birth, or adoption, and thus excludes cohabiting partners as well as unrelated people under age 15.[8] We instead include all people in the household regardless of relationship status, under the assumption that all household members pool resources to meet expenses.

We begin our analysis using the full CPS ASEC to align our estimates with the extant survey-based literature, and then we subsequently impose the following sample restrictions: the individual must be linked to the tax data, and they cannot have their earnings imputed. Individuals in the ASEC are assigned a unique, protected identification key (PIK), and that PIK is used to link the ASEC to tax data (Wagner and Layne, 2014). A household is included in the analysis if there is tax information linked to an adult in the household.

A PIK may be missing if the individual has no history of tax filing or has not provided sufficient personal information to probabilistically predict a PIK. As depicted in Figure C1 (appendix is available online), the link rate averaged around 90 percent of the weighted sample population, though it has trended slightly downward since peaking in 2010. Imputation of earnings occurs because of failure to respond to earnings questions (item nonresponse), or failure to respond to any or enough questions on the ASEC (survey nonresponse). The Census Bureau imputes missing earnings for item nonresponders and imputes the entire ASEC supplement to those who fail to respond to the ASEC (called whole imputes). Figure C1 also shows that the share of people reporting earnings (combined item and survey responders) fell from 80 to 70 percent over our sample period. To retain population

[7] The internal ASEC differs from the public-use version 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.

[8] The typical minor excluded from the official poverty universe is a foster child. The ASEC does not ask income questions to those under age 15, and thus assignment of income to unrelated minors such as foster children requires assumptions on resource-sharing mechanisms within the household, which the Census Bureau avoids with their official poverty universe.

representativeness, we reweight the restricted sample using IPWs. Specifically, for each year and gender, we estimate a probit model of the probability that the individual has a PIK and no imputed earnings as a rich function of demographics, and then divide the person-level ASEC supplement weight by the fitted probability from the regression.[9] The Data Appendix provides further details on sample construction and linking.

As mentioned, our benchmark is the EITC recipient file, which the IRS provides to the Census Bureau for each tax year beginning in 2005. Upon linking the recipient file to the CPS ASEC and appropriately reweighting for selection on PIK placement and nonimputed earnings and income, the matched data cover between 90 and 95 percent of recipients and 92–97 percent dollars paid each year, as confirmed by comparison with internal IRS records.[10] This close correspondence between IRS reports of EITC receipt and our weighted matched data provide the foundation for our estimates of the impact of actual EITC dollars on poverty.

B. Measuring After-Tax and Transfer Poverty

Our focal variable is household after-tax and in-kind transfer income, defined as the sum of labor earnings; rent, interest, and dividend income; cash payments from private individuals and governments, inclusive of welfare and social insurance; and in-kind transfers from food and energy assistance programs; and minus tax payments from federal, state, and payroll taxes inclusive of refundable credits such as the EITC.[11] The labor, nonlabor, and in-kind transfer data come from self-reports in the ASEC survey, while tax payments and the EITC must be simulated or obtained from IRS administrative tax records. Because of its public availability and widespread use across multiple data sets and applications, our default tax simulator is NBER's TAXSIM, version 27.[12] This simulator uses up to 27 input variables

[9] Wooldridge (2007) discusses the advantages and limitations of IPW as a method to address missing data, noting that conditional on covariates it relies on the missing at random assumption. Bollinger et al. (2019) show that earnings nonresponse in the ASEC is 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, Bollinger, and Ziliak (2015) show that imputation in the 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.

[10] Correspondence with Dean Plueger, IRS RICS, January 31, 2022.

[11] The Census Bureau stopped producing a simulated market-value of in-kind housing assistance with the 2016 survey (2015 calendar year), and thus to keep the composition of the household income measure constant over time, we do not include housing benefits since it is missing in the last two years of the survey sample.

[12] We use version 27 of TAXSIM that has been installed for research purposes with the internal ASEC at the Census Bureau. It is the same as Internet TAXSIM (v27) available at https://users.nber.org/~taxsim/taxsim27/. A copy of the code to prep the ASEC can be found at https://sites.google.com/site/jamesziliak/Home/Research.

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, and other taxable and nontaxable income, and potentially deductible expenses (home mortgage interest, property tax). Most of the labor and nonlabor income variables in the ASEC are recorded at the individual level and must first be assigned to tax units before running them through TAXSIM. We modify a program first developed by Judith Scott-Clayton and hosted at the NBER's TAXSIM website to create tax units based on household, family, and relationship variables, altered to more precisely capture features of tax units that feed into EITC eligibility and the updated version of TAXSIM.[13]

To estimate the antipoverty effects of the EITC, we compare the number of people with after-tax and transfer household income below the household-size specific SPM threshold when we include and exclude the EITC from the measure of after-tax income.[14] We do this comparison not only using the *TAXSIM* estimate of the EITC but also swapping out the *TAXSIM* value alternatively with the EITC produced by the in-house *CPS* tax model as well as the actual EITC payments as reported by the IRS.[15] We denote actual EITC payments as *IRS Paid* and use this as our benchmark estimate of the antipoverty effects of the EITC. 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.

In addition to these two EITC simulators and *IRS Paid*, in some of our analyses we also consider a fourth estimate, which we call *TAXSIM Admin*, whereby we replace the survey-based values of earnings, tax-filing status, and number of qualifying children with their respective administrative values provided by our direct link to the tax data. Table 2 summarizes the four tax estimates, and the Data Appendix contains more detail on the CPS and TAXSIM models.[16] We also discuss in the appendix how the SPM poverty threshold is constructed, and we provide summary

[13] NBER hosts both the original code by Scott-Clayton and the code by Ziliak at https://users.nber.org/~taxsim/to-taxsim/cps. We update tax unit characteristics, such as filing status and number of children, when examining the impact of administrative record on estimates.

[14] The SPM thresholds are produced by the Bureau of Labor Statistics using data on food, clothing, shelter, and utilities from the Consumer Expenditure Survey. Details on their construction are available at https://www.bls.gov/pir/spmhome.htm, and in the Data Appendix.

[15] For simplicity, we italicize *CPS* and *TAXSIM* to refer to the estimated EITC payment derived from each simulator. 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, both the data and user-constructed tax units.

[16] An earlier version of this paper also used the tax simulator produced by Bakija (2009). The estimates from his simulator differed little from the CPS and TAXSIM estimates and thus are omitted for ease of presentation.

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 ASEC tax heads
TAXSIM	Input file of 27 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 in the public domain.	All ASEC tax heads
TAXSIM Admin	Same as *TAXSIM*, except using administrative records values of income, earnings, filing status, and qualifying children in place of survey values when available.	All ASEC tax heads
IRS Paid	The EITC recipient file. A CPS ASEC family head must receive a unique identifier and appear in the recipient file to be considered *IRS Paid*.	All ASEC tax heads who receive a unique identifier

statistics for the samples used in the analysis. All income amounts are in real 2016 dollars using the Personal Consumption Expenditure Deflator.[17]

IV. RESULTS

We begin our analysis by providing estimates of the antipoverty effects of the EITC using the full ASEC sample, including those with imputed earnings and those with no link to IRS tax records. This full sample is used by the Census Bureau for its estimates of SPM poverty, and by the extant literature for producing estimates of the effect of the EITC on poverty. These estimates motivate the solutions we introduce as our analysis proceeds, as they highlight several factors that impact differences in linked administrative and survey estimates. Next, we demonstrate our suggested improvements by showing estimates using the restricted sample of those with (1) a link to the tax data and (2) who do not have imputed earnings or the entire ASEC supplement imputed (we examine each restriction separately

[17] Table B-3 in the 2021 Economic Report of the President, https://www.govinfo.gov/content/pkg/ERP-2021/pdf/ERP-2021-table5.pdf.

and the two combined). We then examine the distribution of EITC payments, over all households and by number of qualifying children and household tax-filing status, and how the distribution changes as we substitute survey-based information on earnings, filing status, and children with those provided by administrative data.

A. Antipoverty Effects of the EITC

From the full ASEC, the left panel of Figure 1 provides a comparison between the weighted number of people lifted from after-tax and transfer poverty by each survey-based EITC estimate and *IRS Paid* payments (using Census-provided weights). The *TAXSIM* and *CPS* model estimates suggest that the number lifted out of poverty by the EITC is countercyclical — increasing from just more than four million people in the mid-2000s, then rising to about 5.5 million during the years surrounding the Great Recession, only to fall back to just more than four million by 2016, coinciding with the economic recovery. Regardless of year, around 11 percent of all those in poverty are lifted from below the threshold to above, indicating that the credit's reach increases in tandem with the expanding pool of low-income workers during the recession, but its impact does not change in percentage terms. This suggests that while the EITC provides implicit income insurance to

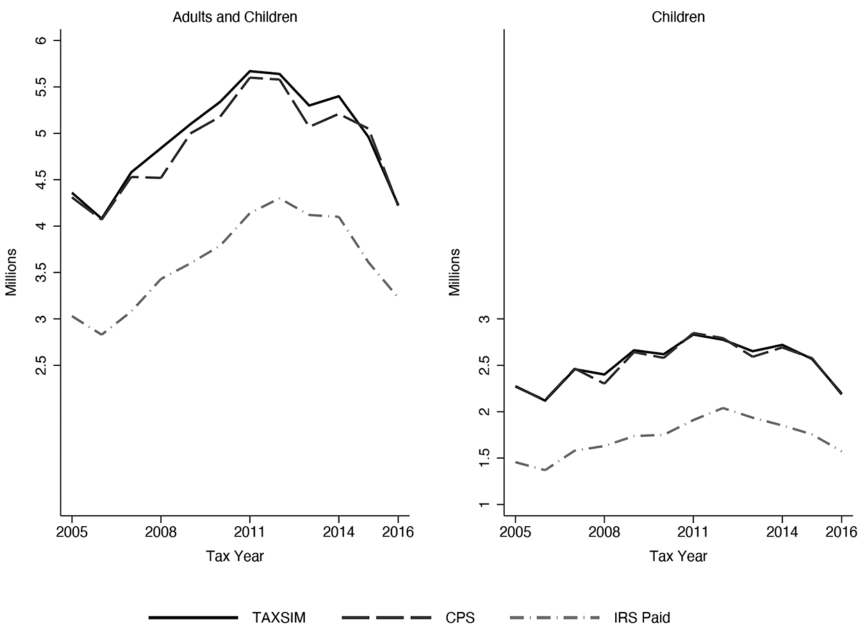

Figure 1. Estimates of the number of people/children lifted from SPM poverty by the EITC using the full ASEC sample and original survey weights. Data are from combined EITC/CP0927 recipient files, CPS ASEC, Form 1040, and Form W-2, tax year 2005–2016. Release authorization number CBDRB-FY21-317.

families in periods of economic distress, serving as part of the system of automatic fiscal stabilizers among those in work, it is not as buoyant as programs like Supplemental Nutrition Assistance Program (SNAP) in the face of economic recessions (Auerbach and Feenberg, 2000; Kniesner and Ziliak, 2002; Ziliak, 2015a; Bitler, Hoynes, and Kuka, 2017).

Although the survey-derived values show similar year-by-year trends, and lie close to one another, in a typical year they result in estimates of the antipoverty alleviation of the EITC about 1.25–1.35 million individuals higher compared with the *IRS Paid* amounts, or at least 35 percent. The right panel shows the same exercise for children, with survey-derived estimates of about 800,000 more children (47 percent) lifted out of posttax and transfer poverty by the EITC than suggested by actual payments made by the IRS. These differences between survey-based estimates and estimates including *IRS Paid* of the antipoverty effectiveness of the EITC are large and fairly persistent over the sample period, and thus we next explore the possible sources of these discrepancies.

The first two sources of the estimation gap we explore may be due to individuals in the CPS ASEC who cannot be linked to the IRS data because they did not receive a PIK, and those who did not report their earnings and thus have those values imputed by Census. Next, we examine how the survey-derived estimates of the EITC and *IRS Paid* affect poverty when we restrict the sample to those ASEC households that included a tax filer who was linkable to the tax data and did not have any survey earnings information imputed, using the IPW procedure described earlier. Failing to adjust survey weights for non-PIKing would mechanically lead to fewer individuals lifted out of poverty by administrative EITC payments. On the other hand, keeping imputed survey earnings would have ambiguous effects depending on the direction in which earnings are imputed and the distribution of imputed and nonimputed earnings. Figure 2 shows the results of this exercise.

The trend line for *IRS Paid* remains largely the same in Figure 2 compared with Figure 1, as would be expected, while the total number of people lifted from poverty via the survey-based estimates drops to align more closely with actual payments. We find that the discrepancies between administrative and tax simulator antipoverty estimates of the EITC are largely eliminated, differing by only about 5–8 percent in a typical year for all people, and by only 1–3 percent in the years after the Great Recession. The gap of 10–15 percent among children is slightly higher than the overall population but only two-thirds of that found in the full ASEC sample depicted in Figure 1. Based on *IRS Paid*, we estimate that 3.9 million people and 1.7 million children are lifted from poverty in an average year over the 2005–2016 sample period. The corresponding estimates from *TAXSIM* and *CPS* are 4.2 and 4.1 million people, respectively, and 1.9 million children. Notably, dropping those without a PIK or with imputed incomes and reweighting only affects the level of those lifted from poverty by the EITC and not the trend, suggesting the EITC offers implicit income insurance over the business cycle as indicated by the strong increase in numbers lifted from poverty during the Great Recession.

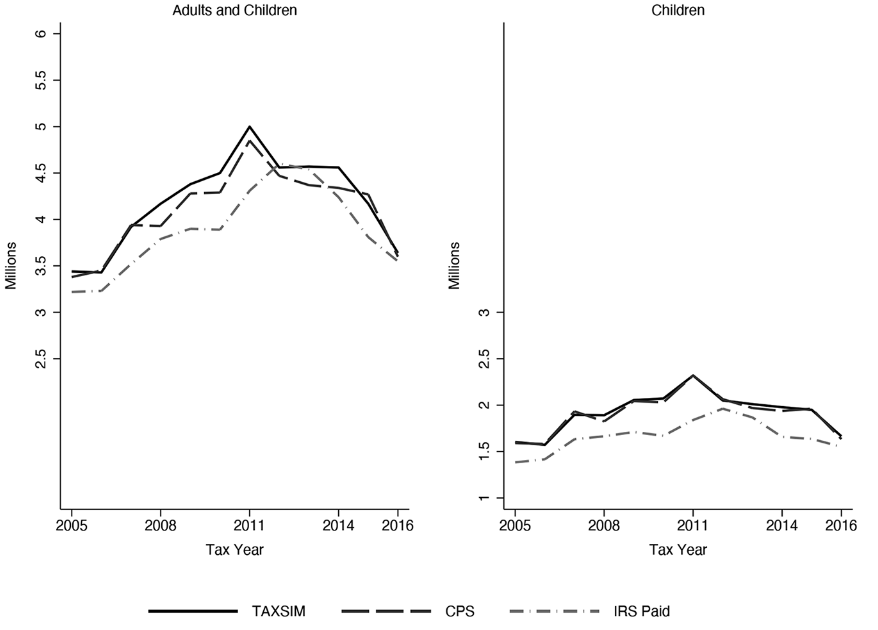

Figure 2. Estimates of the number of people lifted from SPM poverty by the EITC using the sample and weights adjusted for link between the ASEC and IRS data and response to earnings and income questions in the ASEC. Data are from combined EITC/CP0927 recipient files, CPS ASEC, Form 1040, and Form W-2, tax years 2005–2016. Release authorization number CBDRB-FY21-317.

Whether the discrepancy between the survey and administrative estimates of the EITC's antipoverty effects stems from the failure to link the ASEC to tax information (i.e., the absence of a PIK) or the presence of imputed earnings is shown in Figure 3. The figure presents the antipoverty effect for tax units with a PIK, regardless of imputation status, and the effect for tax units without imputed information, regardless of whether they have a PIK. In each case, we predict the probability of being in sample (i.e., with a PIK, or with no imputed earnings) using the same set of covariates as the prior probit IPW model and reweight the person weights accordingly. For ease of presentation, we only show the results for the TAXSIM model and IRS Paid, as those from the CPS tax model are little different. Figure 3 suggests that reweighting based on having a PIK makes a considerable contribution to closing the difference in survey and administrative estimates relative to IPWs based on having no imputed earnings. This does not imply that correcting for nonresponse is less important than correcting for a missing link to tax data. As shown by Bollinger et al. (2019), this result is likely because, conditional on demographics, failing to link is a missing-at-random data problem that IPW can correct. However, earnings nonresponse is nonrandom, even controlling for demographics, which is partially but not fully corrected by IPW.

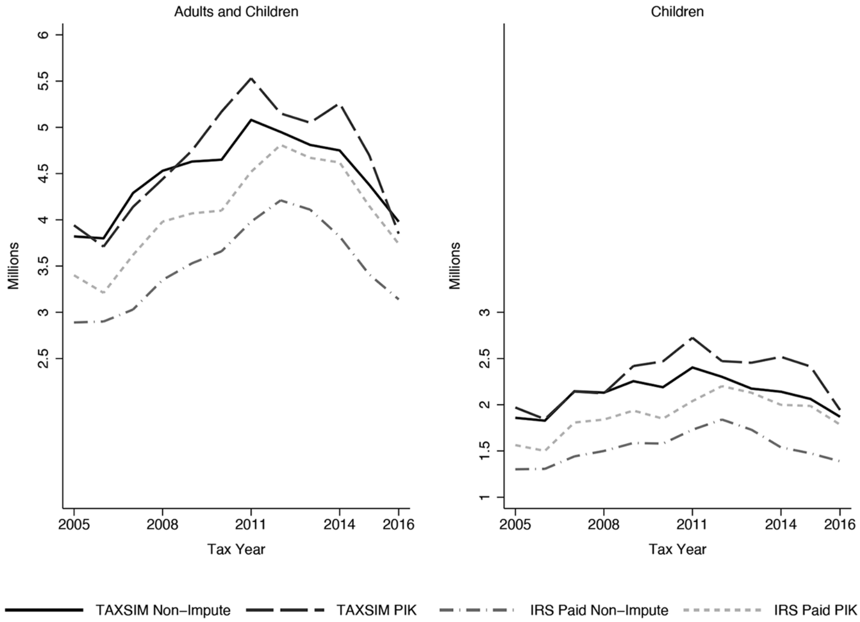

Figure 3. Estimates of the number of people lifted from SPM poverty by the EITC using the sample and weights adjusted for the PIK placement in the ASEC, but not earnings imputation status, and adjusted for earnings imputation status, but not PIK placement. Data are from combined EITC/CP0927 recipient files, CPS ASEC, Form 1040, and Form W-2, tax years 2005–2016. Release authorization number CBDRB-FY21-317.

A possible concern with dropping those without a PIK and reweighting is that the method might understate the number of people lifted from poverty by the EITC. Indeed, Table C.2 suggests that the non-PIKed are socioeconomically disadvantaged relative to those PIKed and more likely to be eligible for the EITC. Because we know PIK status in the restricted-access linked data set, we can construct worst-case bounds of how many we might be undercounting. Specifically, we retain those sample members without a PIK and who have no missing earnings or incomes, and then reweight the ASEC supplement weight by the inverse probability of not having imputed income as this makes the population of interest the non-PIKed.[18] We then estimate the number lifted out of poverty using the TAXSIM and CPS tax simulators, as depicted in Figure 4. The figure shows that had we received a PIK for the non-PIKed, we would have estimated, at most, about 10 percent more people and children lifted out of poverty by the EITC (about 445,000 people; 166,000 children). However, we emphasize that this is a worst-case upper bound, for the following reasons: the non-PIKed are much less likely to be citizens and in possession

[18] We do not reweight for non-PIKing, as this would lead to double counting of the impact of non-PIKed on our estimates.

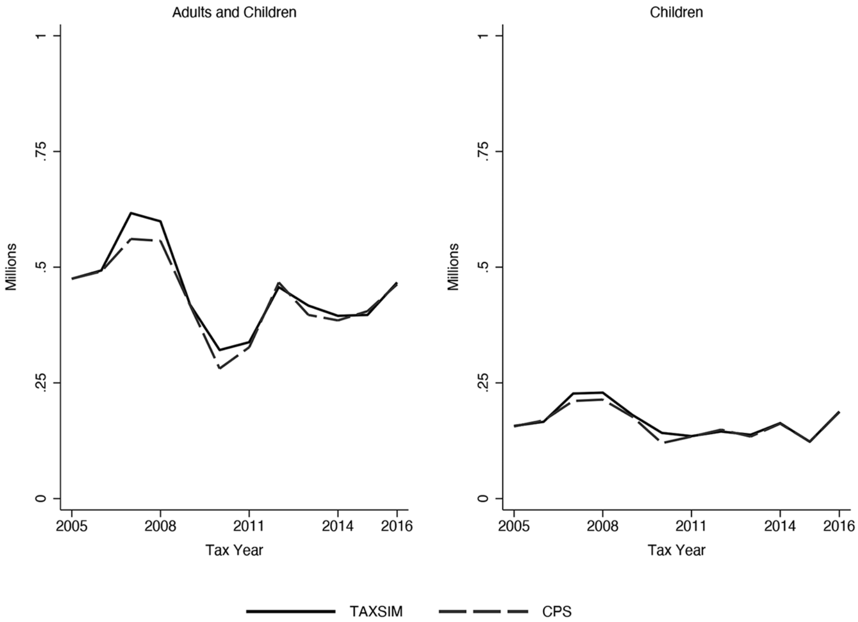

Figure 4. Estimates of the number of non-PIKed people lifted from SPM poverty by the EITC using the sample and weights adjusted for nonimputation of earnings in the ASEC. Data are from combined EITC/CP0927 recipient files, CPS ASEC, Form 1040, and Form W-2, tax years 2005–2016. Release authorization number CBDRB-FY2022-CES010-009.

of a Social Security number, and thus categorically ineligible for EITC; and the non-PIKed are considerably less likely to have ever filed, and consequently would not receive the EITC.[19]

Beyond failure to link and not reporting survey income, another source of this estimation gap is potentially attributable to the assumption of 100 percent take-up for the survey models. That is, the ASEC does not collect information on who does and does not file for the EITC, and the default assumption of the TAXSIM and CPS tax models is that conditional on income and child eligibility, participation is 100 percent. Figure C.2 provides estimates of take-up rates of the EITC for each year, computed as the ratio of those deemed eligible in the ASEC *and* receiving the credit based on the IRS data to those deemed eligible. On average we estimate that about 80 percent of eligibles receive the credit, consistent with prior estimates (Scholz, 1994; Plueger, 2009; Jones, 2014). While suggestive that EITC take-up

[19] Because 1040s are a key component for the master reference file used to match PIKs to individuals, there is a higher probability that non-PIKed heads of household in the ASEC have not filed a 1040 compared with PIKed.

may account for some of the discrepancy between survey and tax data, we note that there is evidence that about one-fourth of actual EITC payments are made improperly (Marcuss et al., 2014). These payments show up in *IRS Paid* but are not estimated via survey-based tax simulators and thus should lead to an attenuation in the gap between the survey and administrative antipoverty estimates.

We conclude this section with some guidance to public users of the ASEC. Although the Census Bureau releases information on who has earnings or their entire supplement imputed, it does not release information on whether the individual can be linked to tax records via a PIK. However, Bollinger et al. (2019) provide evidence that while the average PIK rate among citizens in their ASEC sample is about 88 percent, the rate among noncitizens of Hispanic ethnicity is half that amount. This low link rate is even anomalous among noncitizens — non-Hispanic noncitizens were linked 78 percent of the time, or 34 percentage points higher than among noncitizen Hispanics.

To test whether we can mimic the lack of a PIK in public data, in Figure 5 we drop noncitizen Hispanics and divide the IPW weight in the remaining sample by the gender- and year-specific share who are citizens or non-Hispanic noncitizens. The figure demonstrates that once we drop Hispanic noncitizens in addition to imputed

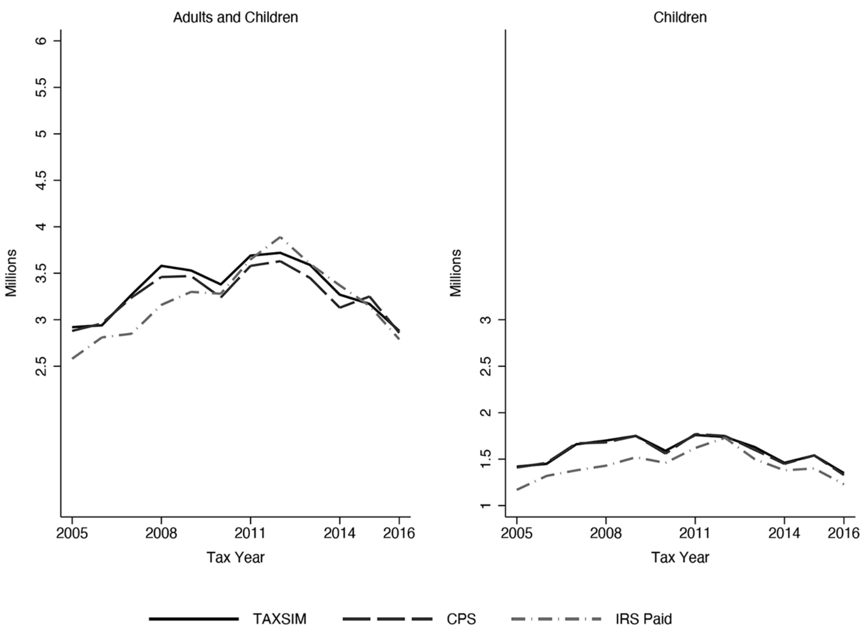

Figure 5. Estimates of the number of people lifted from SPM poverty by the EITC using the sample and weights adjusted for nonimputation of earnings in the ASEC and non-Hispanic citizenship status. Data are from combined EITC/CP0927 recipient files, CPS ASEC, Form 1040, and Form W-2, tax years 2005–2016. Release authorization number CBDRB-FY21-338.

earners, the antipoverty estimates of the EITC are quite similar between the survey-based simulators and tax data. This is especially the case in the years after the Great Recession as we saw in Figure 2 when actual PIK status is known, suggesting that public users of the ASEC can closely approximate the effect of PIK status on EITC poverty estimates using survey information on citizenship and ethnicity.

B. Distribution of EITC Payments

In this section, we delve more deeply into the differences between survey and administrative estimates of the EITC across the income distribution by looking within the household at individual tax-filing units. Because the EITC is applicable to tax units rather than households per se where poverty is measured, any analysis of differences in survey versus administrative values on income, marital status, and number of qualifying children must first take place at the tax-unit level. Unless noted otherwise, the sample used in this section is the restricted sample of those individuals with a link to the tax data and without any imputed earnings, with reweighting using IPW as in Figure 2.

Figure 6 presents the full distribution of the EITC from the pooled 2005–2016 sample over the range of real adjusted gross incomes (AGI) under $60,000, where the preponderance of EITC payments are made. We calculate the average value of the credit over 100 bins of tax-unit AGI, taking three different approaches: first, we bin income using the AGI value as that used in estimating the credit — that is, survey income for *TAXSIM* and *CPS* and administrative income for *IRS Paid* (panel 1); second, we bin both survey and administrative EITC payments over

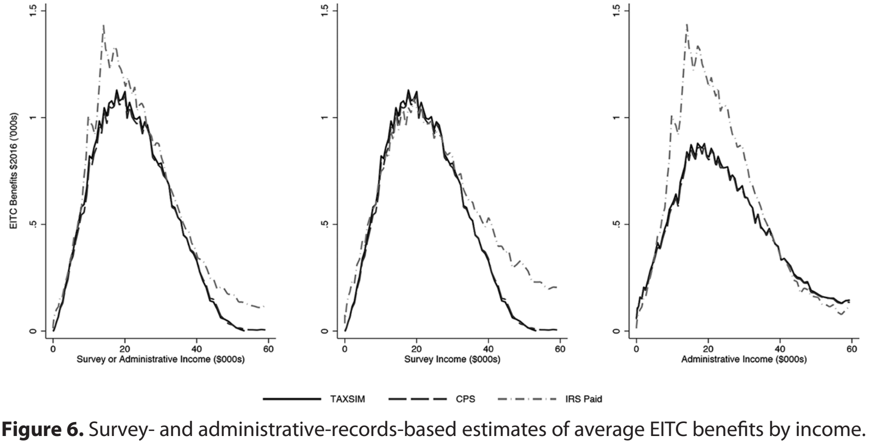

Figure 6. Survey- and administrative-records-based estimates of average EITC benefits by income. Each *x*-axis reports average income within 100 income bins from $0 to $60,000, using the income definition stated in the title. See text for further details. Data are from combined EITC/CP0927 recipient files, CPS ASEC, Form 1040, and Form W-2, tax year 2005–2016. Release authorization number CBDRB-FY21-317.

survey values of AGI alone (panel 2); and third, we bin over administrative values of AGI (panel 3).[20] The second two panels are motivated by prior studies that show that administrative tax data appear to have larger left tails of the earnings distribution than survey data sets (Kornfeld and Bloom, 1999; Abowd and Stinson, 2013; Abraham et al., 2013; Bollinger et al., 2019). This difference could stem from failure of survey respondents to report earnings from part-year (low-earnings) jobs to survey field representatives; or it could result from failure to report earnings to tax authorities. In both cases, the distribution of earnings in tax data is shifted to the left in comparison to survey responses.

Figure 6 shows that the EITC amounts from *IRS Paid* lie above all other estimates over much of the income distribution when it is assessed using administrative income. Meanwhile, survey values of EITC clearly fall to zero at approximately $54,000 when assessed against survey income — as they should — while *IRS Paid* remains strictly positive (though small) above $54,000 regardless of whether survey or administrative income is used along the *x*-axis.[21] Meanwhile, values of *IRS Paid* graphed over administrative income never reach zero, indicating that some actual payments may be paid past the maximum income threshold and thus are possibly erroneous. *IRS Paid* is more "peaked" than the survey estimates at low- to mid-range values of income, consistent with a larger left tail in administrative income, and possible erroneous payments at lower incomes as suggested in the third panel of Figure 6.[22] But this peak at lower incomes also suggests the credit payments are going to financially disadvantaged tax units, and thus on this metric appear to be target-efficient.

Figure 7 breaks out the EITC distributions by number of children claimed and Figure 8 by filing status (note the differing scales across panels). Each figure is constructed in the same manner as panel 1 in the preceding figure, where the income source used for binning is the same as that used to estimate the credit value. We combine information on number of children and filer status 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 survey estimates, in this set of figures we use *TAXSIM* estimates because of its easy accessibility and its

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

[21] The maximum income threshold for a married couple in 2016 is $53,505.

[22] Figure C3 shows the distribution of survey-based EITC payments based on earnings response and PIK status, weighted by the ASEC supplement weights, where the EITC payments for the earnings nonresponders are estimated using Census-imputed earnings values. The Response/No-PIK and Nonresponse/PIK distributions have a higher peak and are shifted leftward relative to the Response/PIK sample used in the figures. This is consistent with failure to PIK and failure to response being more pervasive in the left tail of the earnings distribution, the region where recipients would receive higher EITC payments. The Nonresponse/No-PIK is the group we know least about, and the wider dispersion in the distribution of EITC payments is consistent with challenges of finding a suitable "donor" for imputation.

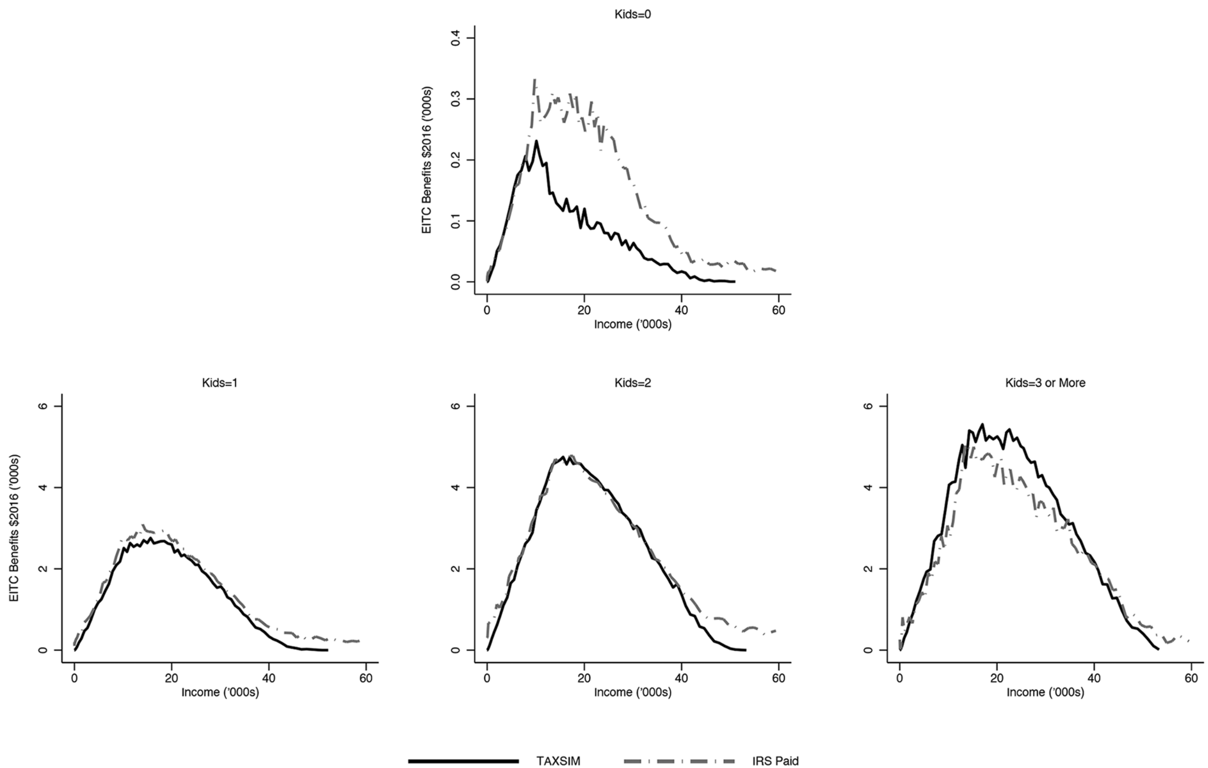

Figure 7. Survey- and administrative-records-based estimates of average EITC benefits by income and number of children claimed. The *x*-axis reports average income within 100 income bins from \$0 to \$60,000, where the income definition is the same one used in estimating the credit value. See text for further details. Data are from combined EITC/CP0927 recipient files, CPS ASEC, Form 1040, and Form W-2, tax year 2005–2016. Release authorization number CBDRB-FY21-317.

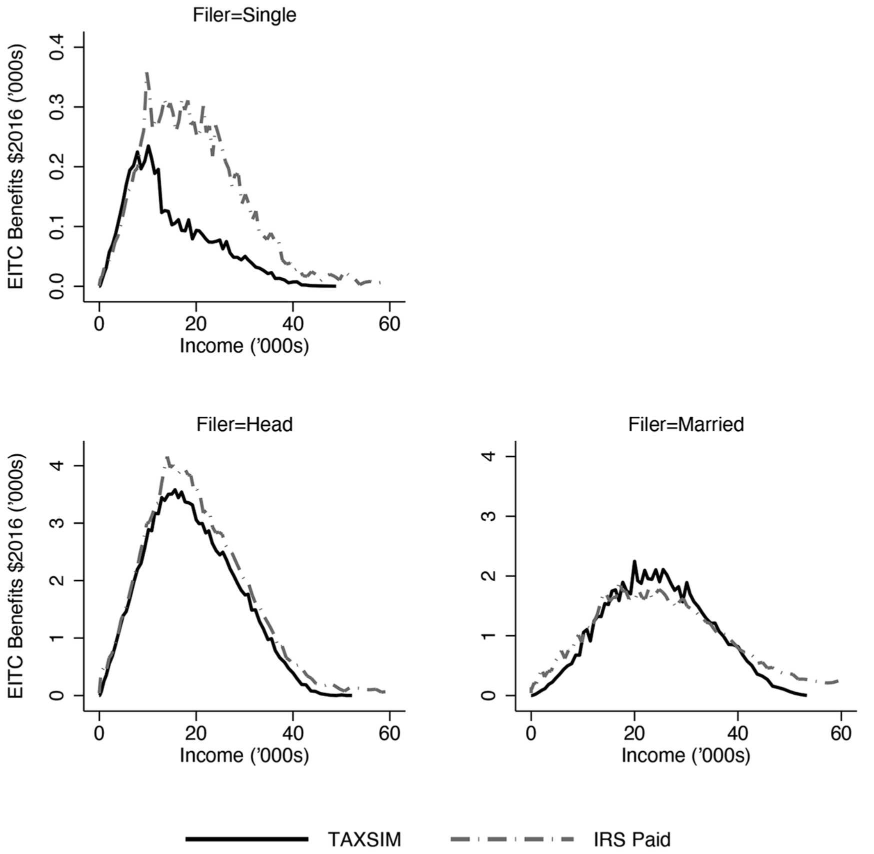

Figure 8. Survey- and administrative-records-based estimates of average EITC benefits by income and filing status. The *x*-axis reports average income within 100 income bins from $0 to $60,000, where the income definition is the same one used in estimating the credit value. See text for further details. Data are from combined EITC/CP0927 recipient files, CPS ASEC, Form 1040, and Form W-2, tax year 2005–2016. Release authorization number CBDRB-FY21-317.

widespread use among researchers. While the *TAXSIM* estimates come close to *IRS Paid* for tax units with one or more children (Figure 7) and for joint filers (Figure 8), the same cannot be said for units without children and with single filers. Absence of children is an obvious source of differences in estimates, but the characterization of unmarried filers is less intuitive. By definition, unmarried filers with children are modeled as heads of household. Thus, due to differences in categorization of unmarried filers based on presence of 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 are claimed by single filers do not account for much of overall EITC expenditure.

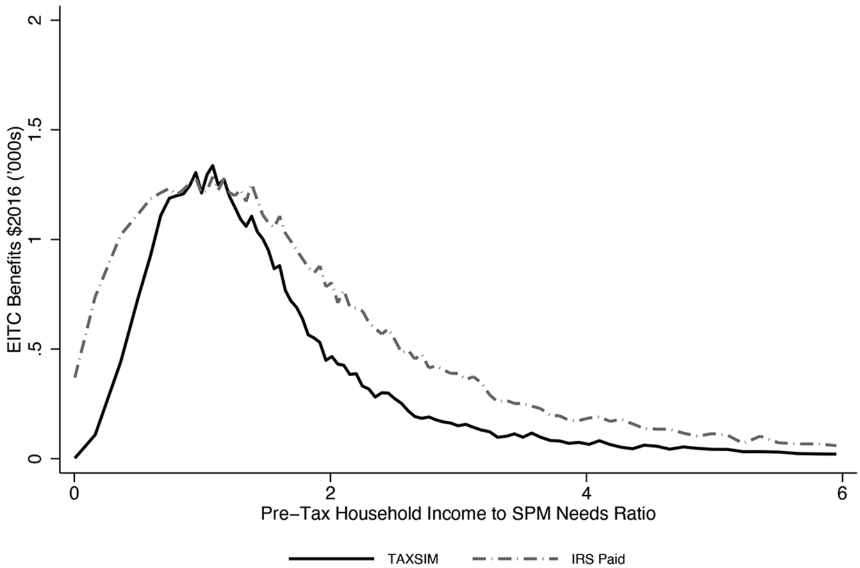

Figure 9. Survey- and administrative-records-based estimates of average EITC benefits for families by the ratio of pretax household income to the household-size-specific SPM threshold. The *x*-axis reports the average ratio within 100 bins of income-to-needs from 0 to 6, where 6 reflects 6 times the SPM poverty threshold. See text for further details. Data are from combined EITC/CP0927 recipient files, CPS ASEC, Form 1040, and Form W-2, tax year 2005–2016. Release authorization number CBDRB-FY21-317.

As a further exploration into whether EITC payments are well targeted to those with low incomes, in Figures 9 and 10 we show the distribution of EITC payments by household income-to-needs. That is, we deflate household pretax income (inclusive of cash and in-kind transfers) by the household-specific SPM threshold, the latter of which adjusts for economies to scale across different household sizes and thus proxies for material needs. A ratio below 1 means the household is in poverty, a ratio below 2 means the household is below twice the poverty line, and so on. The vast majority of payments go to households with incomes less than twice the SPM poverty line, implying that credits are flowing to low-income households as intended by the structure of the credit. As seen in Figure 10, when payments are made higher up in the income distribution, this is primarily among those households with multiple tax filers, some of whom may be eligible for the EITC and some not.

Figures 9 and 10 are particularly instructive in helping resolve a seeming puzzle of Figures 1–3, which show a larger antipoverty effect of the EITC from survey-based estimates compared with *IRS Paid*, and Figure C.4, which shows average EITC payments are higher from *IRS Paid* than either *TAXSIM* or *CPS*. The differences depend on whether a household is close to the SPM threshold based on survey or administrative income. When looking within household, we find that households are about 10 percent more likely to be reclassified as not poor when using their survey income

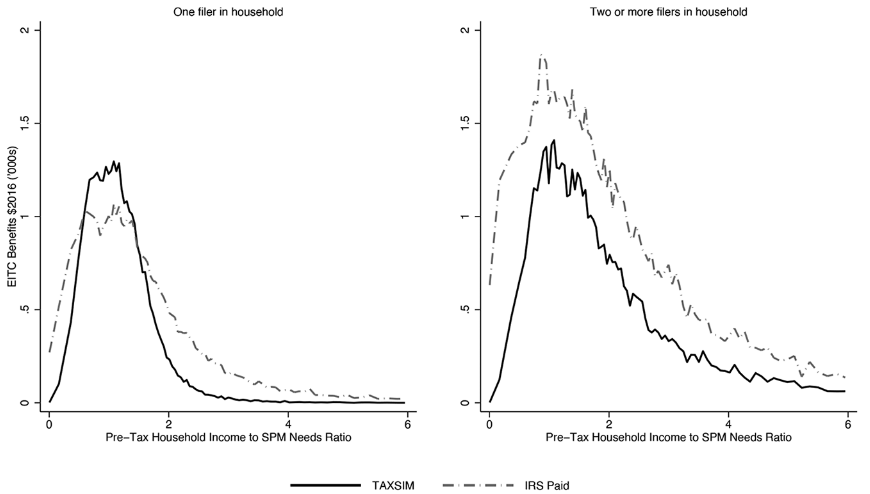

Figure 10. Survey- and administrative-records-based estimates of average EITC benefits for families by the ratio of pretax household income to the household-size-specific SPM threshold and the number of filers in the household. The *x*-axis reports the average ratio within 100 bins of income-to-needs from 0 to 6, where 6 reflects 6 times the SPM poverty threshold. See text for further details. Data are from combined EITC/CP0927 recipient files, CPS ASEC, Form 1040, and Form W-2, tax year 2005–2016. Release authorization number CBDRB-FY21-317.

and EITC estimates than when using their administrative income and *IRS Paid*. Moreover, Figure 9 shows that there is greater dispersion in the *IRS Paid* payments relative to pretax household income-to-needs. The fatter right tail pulls up the average benefit, but the distribution of survey-based estimates is more concentrated around the income-to-needs ratio of 1, which is the antipoverty effect we estimate. The left panel of Figure 10 is for single tax-filer unit households, which are the vast majority of households. Again, one sees greater concentration near income-to-needs of 1 for *TAXSIM*, and it has a higher peak, reinforcing the antipoverty effect from the survey estimates. The right panel shows that the distribution of *IRS Paid* lies everywhere above the *TAXSIM* distribution in multifiler units. This will again pull up the average EITC benefit in *IRS Paid* vis-à-vis *TAXSIM* but will not have a substantive effect on antipoverty effects because the multifilers are few in number.

C. Decomposing the EITC Distribution with Tax Data and Antipoverty Effects

Our next step is to tie our antipoverty estimates with our analysis of how variation in tax characteristics across families influences the calculation of both EITC and poverty. Our goal is to determine the extent to which differences between survey and administrative values on earnings, AGI, filing status, and claimed dependents drive patterns in EITC payments. To make this determination, we make

Figure 11. Survey- and administrative-records-based estimates of average EITC benefits by income, where values for filing status, earnings, and children are iteratively replaced with administrative values. The *x*-axis reports average income within 100 income bins from $0 to $60,000, where the administrative income definition is one used in estimating the credit value. See text for further details. Data are from combined EITC/CP0927 recipient files, CPS ASEC, Form 1040, and Form W-2, tax year 2005–2016. Release authorization number CBDRB-FY21-317.

iterative replacements of administrative values for survey responses and run TAXSIM on the data after each replacement.[23]

The left-hand panel of Figure 11 shows the results of the exercise when we include all sources of earnings. We use bins of administrative income to graph each iteration because full administrative replacement is the goal, except for the line labeled "Survey," which is the original *TAXSIM* estimate from CPS income seen in previous graphs. As we replace survey information, the gap between the survey distribution and the *IRS Paid* distribution shrinks, but it never fully closes. Using administrative filing status in lieu of survey marital status alone has no impact, while using administrative earnings gets us halfway to closing the gap between the two distributions. However, changing both earnings and the number of children claimed nearly closes the entire gap between the two distributions; the only exception occurs over the range of incomes between about $8,000 and $20,000. The right-hand panel provides evidence on why the survey and administrative values

[23] Ideally, we would replace all components of income with administrative values to correct measurement error in SPM income. Since this is not possible using our existing data, we acknowledge that any correction to baseline income may be incomplete, or at least subject to holding all other income inputs constant. If the value of *IRS Paid* is correlated with these other inputs, this would have unknown impacts on our estimates. That being said, our results are similar whether we use SPM thresholds or official poverty thresholds.

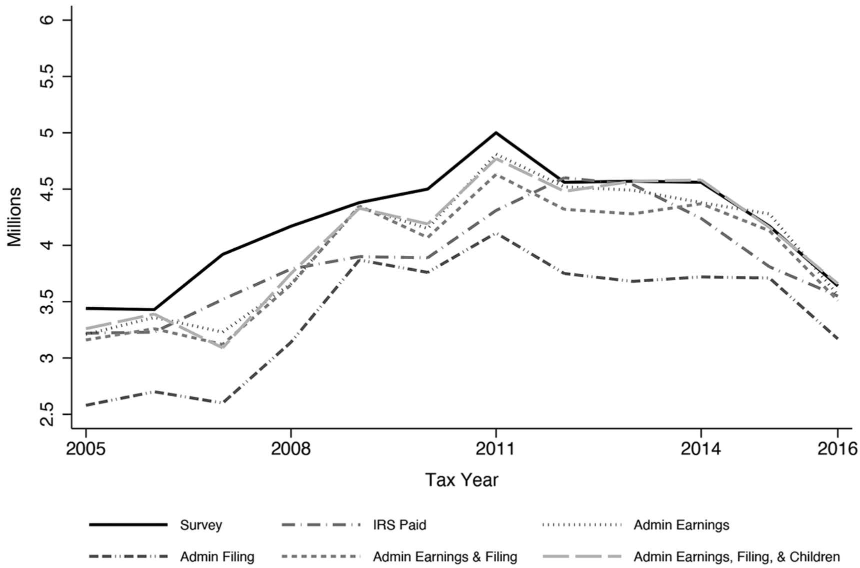

Figure 12. Estimates of the number of people lifted from SPM poverty by the EITC where values for filing status, earnings, and children are iteratively replaced with administrative values. See text for further details. Data are from combined EITC/CP0927 recipient files, CPS ASEC, Form 1040, and Form W-2, tax year 2005–2016. Release authorization number CBDRB-FY21-317.

lead to different results in this range: the influence of self-employment earnings that are reported on a 1040 but are not reported in the ASEC. When we drop observations who report self-employment on a 1040 in the right panel, our swapping in of administrative values for earnings and children leads to a near match in average EITC receipt over the eligible income distribution.[24] The large peak in the lower tail of the EITC payment distribution in the left panel is largely eliminated in the right panel with self-employed workers removed, consistent with research that suggests that the self-employed may manipulate reported income to maximize the EITC (LaLumia, 2009; Chetty, Freidman, and Saez, 2013; Kuka, 2014; Buhlmann, Elsner, and Peichl, 2018).

Our last exercise is to then use the sequential replacement of survey values with administrative values for filing status, earnings, and qualifying children for our anti-poverty estimates of the EITC. Figure 12 presents the results of this decomposition exercise, with the "Survey" and *IRS Paid* series being the same as those reported in the left panel of Figure 2. With the exception of the administrative filing series alone, the other administrative series is generally close to the *IRS Paid* estimates

[24] We also swap out filing status, but as shown in the figures, using administrative filing status has no improvement in the EITC distribution, suggesting that we approximate it well in our survey simulators.e

and, importantly, not too different from the "Survey" series, adding confidence to our survey-based estimates of the antipoverty effects of the EITC.

V. CONCLUSION

We used unique linked survey and administrative tax data to assess the coverage and antipoverty effects of the EITC, showing that about four million people are lifted from poverty in an average year but that the EITC lifts many more during economic downturns like the Great Recession. We compared the antipoverty effects of the EITC using survey-based tax simulators in comparison to actual EITC payments from the IRS. We found that when using the full CPS ASEC, the antipoverty effects of the EITC are much higher when compared with linked actual benefit payments — an estimation gap that is largely driven by CPS ASEC individuals who are not linkable to tax records and who do not report their earnings and incomes to the survey. However, removal of observations with Census-imputed earnings and those without a link to tax records resulted in comparable antipoverty estimates using commonly deployed tax simulators in survey data against administrative records. We also found that actual EITC payments are generally target-efficient at both the tax unit and household levels and that higher actual EITC payments to potentially ineligible households balanced out, in aggregate, the assumption of 100 percent take-up in our survey-based estimates. Considering where EITC recipients land in either the survey or administrative income distributions, it is clear that even potentially erroneously paid EITC dollars are supplementing the incomes of low-earning tax units.

Our research has implications for assessing proposed expansions of the EITC. Although we show that the EITC is countercyclical through its expanded numbers lifted from poverty during the Great Recession, we also find that its antipoverty impact remains flat in terms of the percent of persons lifted from poverty year to year. Many of those lifted from poverty who currently look ineligible based on survey characteristics may become eligible under EITC expansions, which policymakers should take into account when projecting any increases in impact or expected costs of expansion. Finally, researchers using survey data should routinely account for the overstatement of the EITC's antipoverty effects when they assume 100 percent take-up.

Our takeaway recommendation is that public users of the CPS ASEC drop observations with imputed earnings and whole imputes and at a minimum reweight the sample when conducting distributional research on the EITC. Unfortunately, it is not possible to identify those ASEC respondents without a link to tax data in the public domain. However, we demonstrated that it is possible to closely replicate the antipoverty estimates of the EITC in tax data if, in addition to dropping earnings nonresponders and whole imputes, researchers also drop those individuals who are least likely to be linked based on observable characteristics of citizenship and ethnicity and appropriately reweight. We note that for many purposes retaining these households is important for poverty analysis given that they tend to fall in the left tail of the income distribution, and thus removing them should be context specific.

ACKNOWLEDGMENTS AND DISCLAIMERS

We thank the editors, three anonymous reviewers, and comments from Hilary Hoynes, Bruce Meyer, Jon Rothbaum, and seminar participants at the Association of Public Policy and Management Fall Research Conference, Society of Labor Economists Annual Conference, the US Census Bureau, Syracuse University, University of Illinois at Chicago, University of Kentucky, and University of Nevada–Las Vegas. We are also grateful for the assistance of Josh Mitchell on the internal Census files with TAXSIM and to Jon Bakija for sharing his tax simulator used in earlier versions of this paper. Any opinions and conclusions expressed herein are those of the authors and do not necessarily reflect the views of the US Census Bureau. The statistical summaries reported in this paper have been cleared by the Census Bureau's Disclosure Review Board release authorization numbers CBDRB-FY21-317, CBDRB-FY21-338, and CBDRB-FY2022-CES010-009. Research was performed under agreement TIRSE-14-M-00002 between the US Census Bureau and the Internal Revenue Service.

DISCLOSURES

The authors have no financial arrangements that might give rise to conflicts of interest with respect to the research reported in this paper.

REFERENCES

Abowd, John M., and Martha H. Stinson, 2013. "Estimating Measurement Error in Annual Job Earnings: A Comparison of Survey and Administrative Data." *Review of Economics and Statistics* 95 (5), 1451–1467.

Abraham, Katharine G., John C. Haltiwanger, Kristin Sandusky, and James Spletzer, 2013. "Exploring Differences in Employment between Household and Establishment Data." *Journal of Labor Economics* 31 (S1), 129–172.

Auerbach, Alan, and Daniel Feenberg, 2000. "The Significance of Federal Taxes as Automatic Stabilizers." *Journal of Economic Perspectives* 14 (3), 37–56.

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

Bastian, Jacob, and Maggie R. Jones, 2021. "Do EITC Expansions Pay for Themselves? Effects on Tax Revenue and Government Transfers." *Journal of Public Economics* 196, Article 104355.

Bastian, Jacob, and Katherine Michelmore, 2018. "The Long-Term Impact of the Earned Income Tax Credit on Children's Education and Employment Outcomes." *Journal of Labor Economics* 36 (4), 1127–1163.

Bitler, Marianne, Hilary Hoynes, and Elira Kuka, 2017. "Do In-Work Credits Serve as a Safety Net?" *Journal of Human Resources* 52 (2), 319–350.

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* 127 (5), 2143–2185.

Buhlmann, Florian, Benjamin Elsner, and Andreas Peichl, 2018. "Tax Refunds and Income Manipulation: Evidence from the EITC." *International Tax and Public Finance* 25 (6), 1490–1518.

Center on Budget and Policy Priorities, 2016. "Policy Basics: The Earned Income Tax Credit." Unpublished Manuscript, Washington, DC.

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

Citro, Constance F., and Robert T. Michael, 1995. *Measuring Poverty: A New Approach*. National Academies Press, Washington, DC.

Dahl, Gordon, and Lance Lochner, 2011. "The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit." *American Economic Review* 102 (5), 1927–1956.

Duncan, Greg, and Suzanne Le Menestrel, eds., 2019. *A Roadmap to Reducing Child Poverty*. National Academies Press, Washington, DC.

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." *Quarterly Journal of Economics* 111 (2), 605–637.

Fox, Liana, 2019. "The Supplemental Poverty Measure: 2018." US Census Bureau, Current Population Reports, Washington, DC.

Grogger, Jeffrey, 2003. "The Effects of Time Limits, the EITC, and Other Policy Changes on Welfare Use, Work, and Income among Female-Headed Families." *Review of Economics and Statistics* 85 (2), 394–408.

Hardy, Bradley, Charles Hokayem, and James P. Ziliak, 2022. "Income Inequality, Race, and the EITC." *National Tax Journal* 75 (1), 149–167.

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 Moffitt, Robert A. (ed.), *Means-Tested Transfer Programs in the United States*, 141–197. NBER and University of Chicago Press, Chicago, IL.

Hoynes, Hilary, 2019. "The Earned Income Tax Credit." *ANNALS of the American Academy of Political and Social Science* 686 (1), 180–203.

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.

Jones, Lauren E., and Katherine Michelmore, 2018. "The Impact of the Earned Income Tax Credit on Household Finances." *Journal of Policy Analysis and Management* 37 (3), 521–545.

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

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.

Kleven, Henrik, 2019. "The EITC and the Extensive Margin: A Reappraisal." NBER Working Paper No. 26405. National Bureau of Economic Research, Cambridge, MA.

Kniesner, Thomas J., and James P. Ziliak, 2002. "Tax Reform and Automatic Stabilization." *American Economic Review* 92 (3), 590–612.

Kornfeld, Robert, and Howard S. Bloom, 1999. "Measuring Program Impacts on Earnings and Employment: Do Unemployment Insurance Wage Reports from Employers Agree with Surveys of Individuals?" *Journal of Labor Economics* 17 (1), 168–197.

Kuka, Elira, 2014. "EITC and the Self-employed: Real or Reporting Effects?" *Public Finance Review* 42 (6), 691–719.

LaLumia, Sara, 2009. "The Earned Income Tax Credit and Reported Self-Employment Income." *National Tax Journal* 62 (2), 191–217.

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.

Marr, Chuck, Emily Horton, and Brendan Duke, 2017. "Brown-Khanna Proposal to Expand EITC Would Raise Incomes of 47 Million Working Households." Center on Budget and Policy Priorities, Washington, DC, https://www.cbpp.org/research/federal-tax/brown-khanna-proposal-to-expand-eitc-would-raise-incomes-of-47-million-working.

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

Meyer, Bruce D., 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.

Moffitt, Robert, and James P. Ziliak, 2019. "Entitlements: Options for Reforming the Social Safety Net in the United States." *ANNALS of the American Academy of Political and Social Science* 686 (1), 8–35.

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 Moffitt, Robert A. (ed.), *Economics of Means-Tested Transfer Programs in the United States*, Vol. 1, 137–218. University of Chicago Press, Chicago, IL.

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.

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

Schanzenbach, Diane W., and Michael R. Strain, 2021. "Employment Effects of the Earned Income Tax Credit: Taking the Long View." *Tax Policy and the Economy* 35, 87–129.

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

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. US Census Bureau, Washington, DC.

Wooldridge, Jeffrey M., 2007. "Inverse Probability Weighted Estimation for General Missing Data Problems." *Journal of Econometrics* 141 (2), 1281–1301.

Ziliak, James P., 2015a. "Recent Developments in Antipoverty Policies in the United States." In Scholz, John Karl, Hyungpyo Moon, and Sang-Hyop Lee (eds.), *Social Policies in an Age of Austerity*, 235–262. Edward Elgar Publishing, Cheltenham, UK.

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.

Online Appendix

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

April 2022

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.

We also remove people 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. Specifically for each gender and year we estimate a probit model of the probability that an individual has a link to the tax records and has no imputed earnings as a function of education attainment, race, ethnicity, marital status, disability status, nativity, household size, home ownership, and interactions of several demographic categories such as age and education, disability status and education, race and education, race and marital status, and Hispanic ethnicity with nativity. We also control for time-invariant state fixed effects. The probit models suggest that age, education, ethnicity, nativity, and disability status are important model predictors. We then divide the ASEC supplement weight by the fitted probability from the probit regression and use this weight in our analyses of the restricted sample.

A. Income and the SPM

Our focal measure of income is after-tax and transfer income, defined as the sum of gross labor earnings, nonlabor income from rent, interest, and dividend income, cash transfers and social insurance, and in-kind transfers such as food assistance from SNAP and the School Meal Program, and energy assistance. We then subtract federal, state, and payroll taxes, with the latter under the assumption that the individual only pays the employee share of payroll taxes. From this measure we then add in the value of the federal EITC to assess the antipoverty effects of the federal credit (controlling for the state EITC for those states with the state-provided tax credit).

Poverty status of the household is determined by comparing household income to the SPM threshold for a given household composition. The SPM threshold differs from the threshold used in official Census poverty estimates in that the SPM is estimated each year as a rolling three-year moving average of out-of-pocket expenditures on food, clothing, shelter, and utilities collected in the Consumer Expenditure Survey, whereas the official threshold was set in the 1960s and has been updated each year for changes in the Consumer Price Index. See https://www.bls.gov/pir/spmhome.htm for details on the SPM thresholds.

We follow convention of the Census Bureau and adopt the SPM threshold that differs not only by the age composition of the household, but also whether the household owns (with or without a mortgage) or rents their home (Fox 2019). We calculate thresholds for households using a three-parameter equivalence scale, where there are child- and adult-specific adjustments as well as an adjustment for single families. Specifically, the three-parameter scale for one or two adults without dependents is $(\text{adults})^{0.5}$; for single parents it is $(\text{adults} + 0.8*\text{first child} + 0.5*\text{additional children})^{0.7}$; and for other household types it is $(\text{adults} + 0.5*\text{children})^{0.7}$. We then compare each household's income to the appropriate threshold for their household type.

B. Tax Simulators

We adopt two survey-based approaches to simulating the EITC. The first modeling strategy, widely used by many researchers, employs the NBER's TAXSIM model, version 27 (Feenberg and Coutts, 1993). This tax simulator uses up to 27 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). TAXSIM uses the Statistics of Income (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 41 outputs provided by TAXSIM, with the federal EITC (v25) the output of focal interest to our project.

The other 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 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).

For *TAXSIM*, we use the ASEC to create tax units and assign dependents to filers, rather than use Census-constructed input variables such as filing status.[1] The first task is to assign the

---

[1] 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 https://users.nber.org/~taxsim/to-taxsim/cps/cps-clayton/.

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, of an unrelated subfamily, or a primary individual. We also allow for dependent filers.[2] 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. 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.

C. Administrative Estimates

Our administrative value of the EITC derives from those tax units in actual receipt of EITC payment by the IRS regardless of eligibility, which we denote as *IRS Paid* in the main text of the paper. We match the EITC recipient file to the survey data using the individually assigned PIK. Appendix Figure C1 shows the link rate for each year of our sample, where the probability of being linked has remained at around 90 percent, with a slight deterioration in more recent years. Meanwhile, the rate of non-imputed responses has deteriorated strongly over time, with rates dropping from 80 percent in 2005 to 70 percent in 2015.

---

[2] In other words, we take any children who filed a 1040 (teenage workers, for example), and assigned them to the tax unit of the person who claimed them as a dependent.

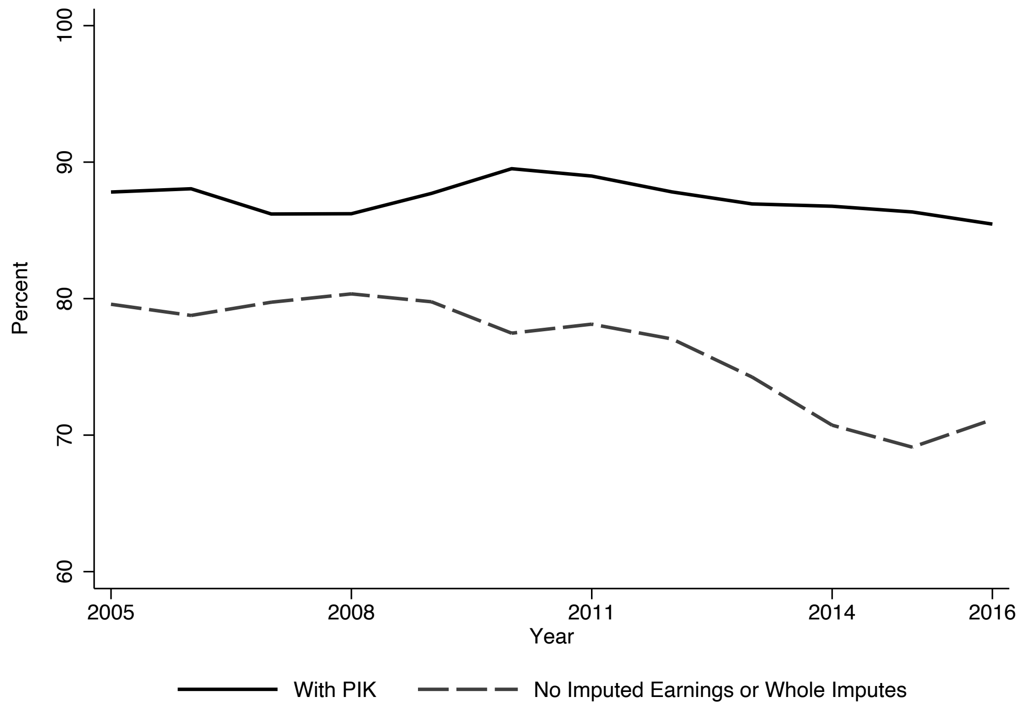

Figure C1. PIK and non-imputation rates over time. Source: combined EITC/CP0927 recipient files, CPS ASEC, Form 1040, and Form W-2, tax years 2005-2016. Release authorization number CBDRB-FY21-317.

Take-up rates of the EITC, defined as the ratio of the number of taxpayers estimated to be eligible and actually receiving payment to the number of estimated eligible, are shown in Figure C2. These are the rates produced for the EITC fact of filing project using the same data as used in this paper. They range between 77 and 81 percent over the period.

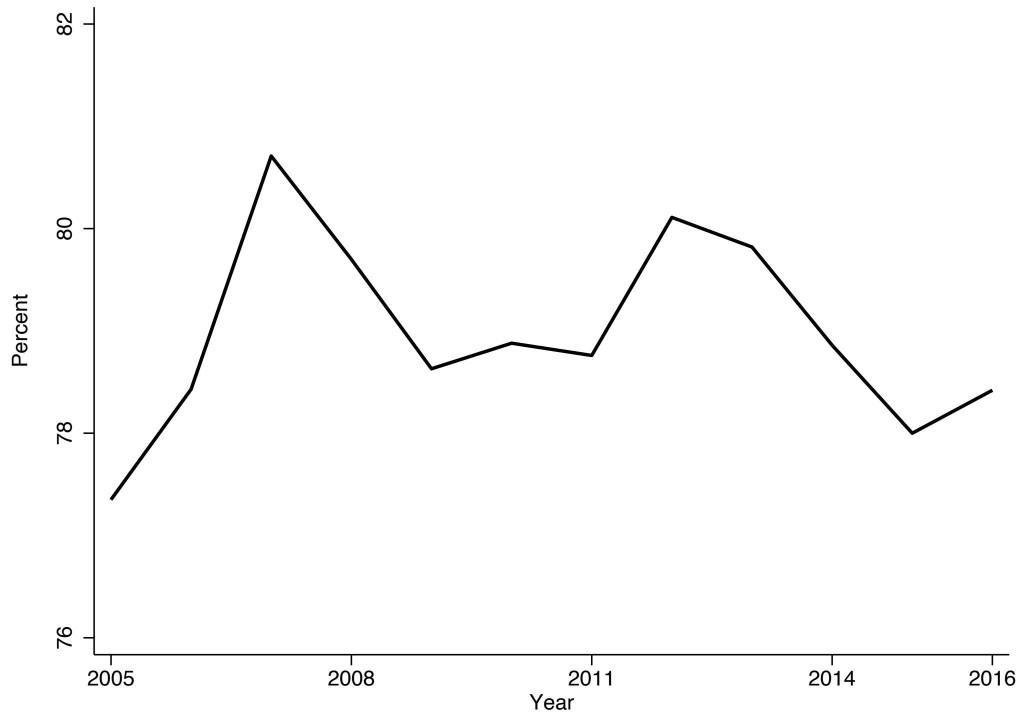

Figure C2. EITC take-up rates over time. Source: combined EITC/CP0927 recipient files, CPS ASEC, Form 1040, and Form W-2, tax years 2005-2016. Release authorization number CBDRB-FY21-317.

Figure C.3 presents the distribution of EITC payments based on earnings imputation status and whether the person is PIKed to tax records. The nonresponse lines utilize Census imputed earnings values.

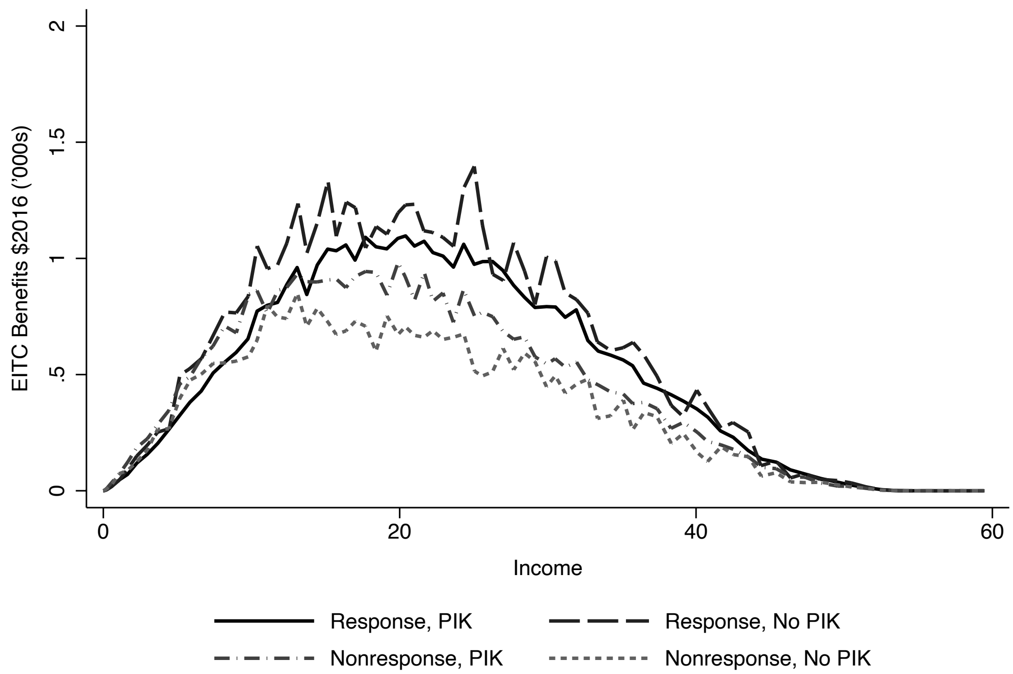

Figure C3. Survey-based estimates of average EITC benefits by income and response category. The *x* axis reports average income within 100 income bins from $0 to $60,000. See text for further details. Source: combined EITC/CP0927 recipient files, CPS ASEC, Form 1040, and Form W-2, tax year 2005-2016. Release authorization number CBDRB-FY21-317.

Table C1 provides summary statistics on the sample, the mean eligibility rates, and the mean values of the estimators. The survey-based estimators have similar mean eligibility values, while the administrative value shows higher receipt rates and dollar values, most noticeably for the sample of PIKed and non-imputed filers.

Table C1.a: Summary statistics for full ASEC sample using person weight

	Mean	SE
EITC Receiver (TAXSIM)	0.13	0.0004
EITC Receiver (CPS)	0.13	0.0003
EITC Receiver (IRS Paid)	0.13	0.0004
EITC Amount (TAXSIM)	267.4	0.935
EITC Amount (CPS)	264.1	0.937
EITC Amount (IRS Paid)	319.2	1.063
Household EITC Amount(TAXSIM)	408.6	1.184
Household EITC Amount (CPS)	402.8	1.198
Household EITC Amount (IRS Paid)	513	1.438

Household Annual Earnings	65440	95.24
Household Nonlabor Income	15140	30
Household After-Tax Income (TAXSIM)	66050	67.21
Age	47.19	0.0203
Male	0.61	0.0005
White	0.79	0.0004
Native Born	0.83	0.0004
Not Hispanic	0.85	0.0004
Not Disabled	0.90	0.0003
Less than HS	0.13	0.0004
High school grad	0.31	0.0005
Some college	0.28	0.0005
College degree	0.29	0.0005
Homeowner	0.64	0.0005
Household Size	2.79	0.0017
Number of Qualifying Children	0.65	0.0011

Table C1.b: Summary statistics for ASEC sample without imputed earnings and linked to tax records, using reweighted person weight

	Mean	SE
EITC Receiver (TAXSIM)	0.13	0.0004
EITC Receiver (CPS)	0.13	0.0004
EITC Receiver (IRS Paid)	0.15	0.0005
EITC Amount (TAXSIM)	284.4	1.291
EITC Amount (TAXSIM Admin Earnings)	302.9	1.311
EITC Amount (TAXSIM Admin Filing)	287.6	1.3
EITC Amount (TAXSIM Admin Earnings and Filing)	312.3	1.32
EITC Amount (TAXSIM Admin Earnings, Filing, and Children)	333.7	1.382
EITC Amount (CPS)	280.5	1.291
EITC Amount (IRS Paid)	375.7	1.516
Household EITC Amount(TAXSIM)	389.4	1.537
Household EITC Amount (CPS)	384	1.554
Household EITC Amount (IRS Paid)	542.9	1.936
Household Annual Earnings	62580	115.6
Household Nonlabor Income	15440	36.69
Household After-Tax Income (TAXSIM)	64050	81.33
Household After-Tax Income (TAXSIM Admin Earnings)	55280	242.6
Household After-Tax Income (TAXSIM Admin Filing)	60220	96.31
Household After-Tax Income (TAXSIM Admin Earnings and Filing)	55750	243.1
Household After-Tax Income (TAXSIM Admin Earnings, Filing, and Children)	55760	243.1
Age	47.61	0.0248

Male	0.62	0.0007
White	0.79	0.0006
Native Born	0.83	0.0005
Not Hispanic	0.85	0.0005
Not Disabled	0.90	0.0004
Less than HS	0.13	0.0004
High school grad	0.31	0.0006
Some college	0.27	0.0006
College degree	0.29	0.0006
Homeowner	0.64	0.0007
Household Size	2.77	0.0022
Number of Qualifying Children	0.67	0.0014

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 2016 dollars using the personal consumption expenditure deflator. Release authorization number CBDRB-FY21-317.

In Figure C.4 we show how well the survey-alone TAXSIM and CPS modeling strategies translate into average inflation-adjusted credit amounts in each year (in thousands of 2016 dollars) compared to IRS Paid. The left panel is for the population overall including zero credit amounts and the right panel for the subsample of EITC recipients (thus note the different scales). The sample used is the restricted sample of individuals with a link to tax data and no imputed earnings or ASEC supplement, reweighted using IPW. The figure shows that the mean values from TAXSIM and CPS 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 when zeros are included, but remaining elevated among the subpopulation of recipients. The former reflects the economic expansion in the last decade where fewer workers were eligible for the credit as the economy recovered. However, both estimators fall short of the actual paid amount from the recipient file (IRS Paid)—by \$80-\$100 depending on year with zeros included and \$200-\$400 among the subset of recipients. Higher average values of IRS Paid do not necessarily translate into higher antipoverty impacts, however, since higher IRS Paid dollars may accrue to tax units whose income already places them above the SPM poverty threshold.

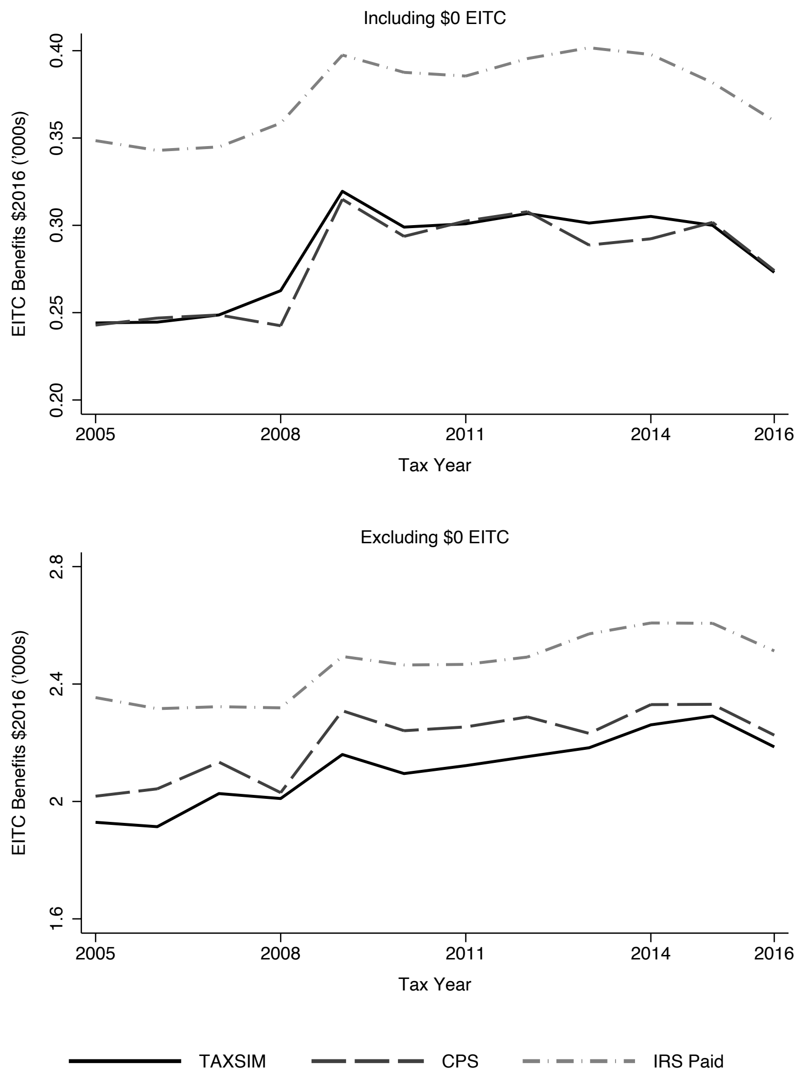

Figure C.4. Mean estimated EITC from TAXSIM, and Census CPS tax models, and paid EITC from combined CPS ASEC and IRS data. Source: combined EITC/CP0927 recipient files, CPS ASEC, Form 1040, and Form W-2, tax years 2005-2016. Release authorization number CBDRB-FY21-317.

Table C.2 presents sample means based on whether the person can be linked to tax records. The table shows that those individuals who cannot be linked have much lower incomes and homeownership rates, have lower education attainment, are less likely to be born in the United States, and more likely to declare Hispanic ethnicity.

Table C.2. Selected Sample Means by PIK Status

	Piked	Not Piked
EITC Receiver	0.19	0.25
EITC Amount	518.2	719.4
Household EITC Amount	2225	2392
Household Annual Earnings	76080	45060
Household Nonlabor Income	13150	7425
Household After-Tax Income	72480	43910
Age	35.68	31.37
Male	0.48	0.49
White	0.79	0.73
Native Born	0.88	0.72
Not Hispanic	0.84	0.70
Not Disabled	0.93	0.94
Less than high school	0.37	0.48
High school graduate	0.21	0.22
Some college	0.21	0.15
College degree	0.21	0.16
Homeowner	0.71	0.55
Household Size	3.52	3.78
Observtions	2,125,000	293,000

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 2016 dollars using the personal consumption expenditure deflator. Release authorization number CBDRB-FY21-338. EITC and tax estimates are based on NBER's TAXSIM.

D. Comparison of internal EITC file to SOI reports

Our comparison of estimates begins with providing an assessment of weighted CPS individuals 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 D1 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 D1. 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.

Additional References

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

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.