CARRA Working Paper Series
Working Paper 2017-10

Tax Preparers, Refund Anticipation Products, and EITC Noncompliance

Maggie R. Jones
U.S. Census Bureau

Center for Administrative Records Research and Applications
U.S. Census Bureau
Washington, D.C. 20233

Paper Issued: December 2017

Tax Preparers, Refund Anticipation Products, and EITC Noncompliance[1]

Maggie R. Jones
U.S. Census Bureau

November 15, 2017

Abstract

Abstract: In this work, I examine whether the availability of tax refund anticipation products (either in the form of a loan or a temporary bank account) is associated with higher non-compliance rates for the Earned Income Tax Credit (EITC). Refund anticipation products are offered by tax preparers as a way for taxpayers to receive a refund faster or to have the tax preparation fee paid from the refund (or both). These products are, on average, costly for taxpayers compared with the average value of a refund, and they are often marketed to low-income taxpayers who may be liquidity constrained or unbanked. Both tax preparers and taxpayers have perverse incentives to use these products, and the temptation of a large refund (for the taxpayer) and added fees and interest (for the tax preparer) may induce erroneous claiming of credits. I examine the association between refund anticipation product use and the overpayment of EITC using tax records and survey data linked at the individual level. For taxpayers in the Current Population Survey Annual and Social Economic Supplement, I estimate EITC eligibility based on household characteristics and combined survey and administrative income information; I can observe EITC credit receipt, the use of paid tax preparation or online filing, and the receipt of a refund anticipation product. I find that both the incorrect payment of EITC and the value of EITC overpayment are associated with preparer use, and to a lesser extent with the use of online filing, when compared with paper filing. Incorrect payment is exacerbated for preparer and online filing when a refund anticipation product is purchased. I also exploit an exogenous price shock to the tax preparation industry that occurred in 2010. This allows me to separately identify a "preparer effect" on EITC noncompliance. I find that the rate of incorrect payment and the dollar value of overpayment increased in the tax year of the shock for those using a preparer and buying a product.

Keywords: Tax compliance, EITC

JEL Classification: H26, K23

[1] Disclaimer: This paper is released to inform interested parties of research and to encourage discussion. The views expressed are those of the author and not necessarily those of the U.S. Census Bureau. Research was performed under agreement TIRSE-14-M-00002 between the U.S. Census Bureau and the Internal Revenue Service. Special thanks to seminar participants at UCLA and the Census Bureau, plus Randy Akee, Jim Ziliak, Alan Plumley and Janet Holtzblatt, for constructive feedback and suggestions.

1. Introduction

Tax refunds are an important source of financial support and consumption smoothing for many households. The social safety net of the United States now largely comprises refundable credits that go to low-income tax filers, reflecting a transition from out-of-work to in-work benefits that dates back to welfare reform (Bitler, et al., 2014). Meanwhile, the complexity of the tax system has led to the widespread use of tax preparers, especially among low-income filers who are concerned about receiving needed credits while avoiding an audit (Rothstein & Black, 2015). Preparation services are expensive, especially when the costs of refund anticipation products are added to the tax-preparation fee. Because such products are lucrative for tax preparers, both taxpayers and preparers face perverse financial incentives to incorrectly claim tax credits or understate income on a tax filing to increase the value of a refund.

In this paper, I use a unique data set to evaluate the impact of paid preparers and refund anticipation products on Earned Income Tax Credit (EITC) noncompliance. These data are derived from the tax records and survey data, linked at the individual level, used to create the yearly take-up estimates of the EITC. The U.S. Census Bureau reports these estimates to the Internal Revenue Service (IRS). This annual report also includes an estimate of the incorrect payment rate of EITC. I am able to observe whether an incorrectly paid filer used a tax preparer and whether he or she purchased a refund anticipation product when filing. To disentangle the paid preparer's incentives to incorrectly file from the filer's incentive, I exploit an exogenous shock to the price of refund anticipation loans (RALs), which caused the abrupt withdrawal of tax preparers from offering this product. This shock led preparers to increase sales of a comparable product that had a lower base price, but to which preparers could add costs associated with higher-refund filings (Wu & Best, 2015). I find that this incentive led to higher rates of incorrect payment in the tax-filing season succeeding the shock—both the probability of incorrect payment and in the value of overpayment. Because of the unique nature of the data, to my knowledge this paper represents the first estimates of the direct effect of paid-preparer incentives on income tax noncompliance.

There is strong anecdotal evidence that preparers respond to the perverse incentives created by higher-priced filings, to which extra fees and products can be added.[2] Moreover, tax preparers face no regulation except for some slight oversight at the state level (Levy, 2015). However, it is extremely difficult to separate the response to incentives on the part of the preparer versus the filer. Low-income taxpayers face their own strong incentives to claim the EITC, the Child Tax Credit, and the head of household filing status (Jones & O'Hara, 2016). Taxpayers with the intent to evade may seek out tax-preparation services in the hope that using a professional will help them avoid scrutiny (Book, 2009). At the same time, filings that include many credit claims and a large refund are especially lucrative for tax preparers, who can charge a higher price for filings with multiple forms and added products.

Disentangling the response to these perverse incentives is an important task. The EITC has an overpayment price tag in the tens of billions of dollars each year—a fact that has brought intense criticism to what is otherwise widely considered an extremely effective policy (see Nichols and Rothstein (2015) for a review of assessments of the program).[3] Evaluating the contribution of the tax-preparation industry to the incorrect payment rate will inform policymakers of the best ways to approach reform while protecting eligible recipients.

In the next sections of the paper, I discuss the background of tax-preparation services, online filing, and the growth of refund anticipation products. I also discuss the literature relevant

---

[2] The National Consumer Law Center has multiple "mystery shopper" reports regarding tax preparation and incorrect filing (www.nclc.org).

[3] http://thehill.com/policy/finance/274046-house-republicans-ask-irs-about-tool-to-prevent-improper-tax-credit-payments

to tax preparation and EITC noncompliance on the preparer side and the filer side. I then give a description of the data in section 3, focusing on the estimation of EITC eligibility and ineligibility. I present the econometric model in section 4, describing the exogenous shock and how it is used in the model. Section 5 provides the results of the econometric model, and section 6 provides some further specification analysis in support of the results. Section 7 concludes.

2. Background and previous literature

2.1 History of refund anticipation products

RALs came into existence during the late 1980s in tandem with the increase in electronic filing (e-filing) (Kopczuk & Pop-Eleches, 2007). They are short-term loans of a refund that tax preparers offer and consumer finance companies underwrite. In exchange for receiving refund money more quickly (often on the same day as filing), a taxpayer pays interest and fees on top of any preparation costs already incurred. Tax preparers make substantial profits on taxpayers through the use of these instruments because the loans are secured by the refund, lowering the risk of making the loan (Wu, et al., 2011). With APRs generally greater than three-digits, these products can be extremely lucrative. Meanwhile, a refund anticipation check (RAC) is a temporary bank account for the taxpayer into which his or her refund is deposited. The tax preparer then writes a check or—more often—provides a pre-paid bankcard for the taxpayer's use. The taxpayer can put off paying the tax-preparation fee and have the fee taken from the refund before the balance is deposited, enabling financially constrained taxpayers to have their taxes prepared without having to pay up front. The product is essentially a lending of the tax-preparation fees, which can be substantial (Wu, et al., 2011). If a taxpayer pays $30 to defer payment of a $200 tax preparation fee for 3 weeks, the APR would be equivalent to 260 percent (Wu & Feltner, 2014).

RALs were the original product offered and, until recently, had a higher price than RACs (Wu & Best, 2015). The history of RALs is tied up in the parallel history of e-filing. The IRS introduced e-filing in 1986 (Kopczuk & Pop-Eleches, 2007). E-filing had benefits both for the IRS in terms of lower administrative costs and for the taxpayer in the increased swiftness of refund receipt and the lower error rate on completed returns. The rollout of e-filing happened through a coordinated effort between the IRS and commercial tax preparers. The very first e-filing program occurred in 1985 through an IRS partnership with H&R Block. The involvement of the tax-preparation industry essentially required that taxpayers use tax-preparation services if they wanted to e-file (Davis, 1998).

Beginning in 2000, the IRS reinstated the debt indicator (which it had provided for a brief time in the 1990s). The indicator alerted tax preparers to any liens against a taxpayer's refund before they approved a RAL, which significantly lowered the risk of the loan and made RALs especially profitable. The IRS stopped providing the indicator in 2010, which led to an abrupt drop in preparers' willingness to offer the loans. The impact of the debt indicator removal occurred through the lenders who financed RALs, when the Federal Deposit Insurance Corporation notified the lenders that making the loans without the indicator would be "unsafe and unsound" (Hayashi, 2016). A simultaneous blow to the market happened in the same year, when HSBC, H&R Block's lender, was forced out of the market by the Office of the Comptroller of the Currency. For lenders who still made RALs, the size of the maximum loan they were willing to provide decreased substantially—from $10,000 to between $750 and $1,500, depending on the lender (Hayashi, 2016).

As tax preparers withdrew from the RAL market, the provision and purchase of RACs increased. Historically, RACs were less expensive than RALs, but as RACs have replaced RALs in the market, their price has gone up (bringing their price closer to that of a RAL) (Wu & Best, 2015). Importantly, much of the increase in price is through add-on fees, including document processing, e-filing, and transmissions fees, which are correlated with filings for credits and thus the size of the refund. In 2011, a group of "mystery shoppers" tried to get a sense of the accuracy and cost of tax-preparation services. The tax preparers who were studied charged the highest fees, between $330 to $540, for a tax-preparation/RAC combination for returns with qualifying-child EITC claims (Wu, et al., 2011). This represents a full 20 percent of the EITC for an average recipient household (Rothstein & Black, 2015).

In 2014, the last year of data in my study, 21.6 million taxpayers obtained a RAC. The National Consumer Law Center estimates a baseline price of a RAC at $30, giving a lower bound of $648 million extracted from tax refunds just for the price of the product alone. The addition of estimated add-on fees brings this amount up to $848 million (Wu & Best, 2015).

2.2 Previous literature

Taxpayer advocates point out that filers who use refund anticipation products are borrowing their own money—money which will come to them in a matter of a few days if they simply e-file (Rothstein & Black, 2015). Why, then, are they so widely used? Researchers have found that consumers of these products want to get their refund money sooner and pay off more pressing debt. In a small study of taxpayers in Detroit, Barr and Dokko (2008) found that 73 percent of unbanked users of tax preparers obtained a RAL so they could pay off bills faster. In the same study, half reported that they needed to take out the loan to pay tax-preparation fees. The population of product users overlaps with users of other types of short-term, high-interest loans, including payday and title loans. Specifically, they tend to be young adults from low-income households (Feltner, 2007; Theodos, et al., 2011) who have children and are more likely to file as heads of household (Elliehausen, 2005; Masken, et al., 2009). Most users of the product have few other opportunities to obtain credit; a quarter of customers do not have a bank account or bank credit card (Elliehausen, 2005). In short, users of tax-preparation products are highly liquidity constrained and lack access to traditional banking and credit.

Providers of these products—similar to other providers of high-interest, short-term loans—use the foregoing evidence to argue that these products make consumers better off. RALs and RACs may provide the only way for low-income filers to acquire credit or to get their refund without having to pay a costly up-front fee when there are other bills to pay (Rothstein & Black, 2015). However, the economic literature on the potential benefits of short-term borrowing is mixed. While Zinman (2010) found generally positive effects of opening up credit markets to liquidity-constrained borrowers, Melzer (2011) found that access to payday lending increases a household's difficulty in paying important bills, such as mortgage, rent, and utilities. Jones (2016) found that similar access to RALs was associated with increased hardship.

In an examination of taxpayers in Illinois, Dewees and Parrish (2009) found that 60 percent of RAL users were EITC recipients, and 72 percent of RAL users in neighborhoods with a high African-American population were EITC receivers. The EITC plays an important role in the size of refunds; unlike withholdings, the EITC is refundable and offsets any tax liability (Jones & O'Hara, 2016). From a public policy perspective, much of the money that is captured from taxpayers through tax-preparation products and services are at their source transfers to low-income taxpayers from other, higher-income taxpayers (Eissa & Hoynes, 2008). When tax preparers charge usurious fees on refund anticipation products, they capture a large portion of the safety net that is meant to support low-income wage earners (Wu & Best, 2015).

While consumer advocates object to the usurious nature of these products, a further consideration is the necessity of tax-preparation services. Filing income taxes is a federal legal requirement, and while low-income taxpayers have options to file at no cost, many may be unwilling to risk an audit or forgo needed credits through a misunderstanding of the tax system. The existence of for-profit tax preparation itself may cloud the importance of how complex the tax system is and dampen the saliency of reform (Finklestein, 2007).

An argument can be made that payments to tax preparers reduce the net benefits of the EITC for many taxpayers—especially if they would not seek out preparers in the absence of the credit. For recipients, those compliance costs would be lower if the IRS helped them prepare their returns at no cost (through volunteer tax preparation sites, for example) or determined eligibility without additional information from claimants. However, IRS administrative costs probably would be higher (unless funding on other agency activities were cut).[4] In considering the view of professional tax preparation as a substitute for IRS administrative costs, however, thought should be given to the regressivity of preparer fees and the lack of federal oversight of tax preparers.

These concerns are compounded by studies that indicate the opportunity of capturing public moneys through the tax and transfer system incentivizes fraud (Wu & Feltner, 2014). Masken, et al. (2009) found that taxpayers who used bank products were more often non-compliant than those who did not. When fraud is uncovered, taxpayers often bear the consequences of non-compliance; it is often difficult to prove fault on the part of the preparer, especially in cases in which the preparer has a seasonal establishment or the preparer did not sign the return (as

[4] Email with Janet Holtzblatt, Congressional Budget Office, August 24, 2017.

required by law) (Levy, 2015). When the fault of non-compliance falls on the taxpayer, penalties usually include at least the reimbursement of a refund and, possibly, the denial of eligibility for credits in later tax years (Levy, 2015).

Taxpayers face strong incentives to be noncompliant, and their noncompliance may be associated with tax-preparer use. Many filers believe that using a professional preparer reduces their risk of audit (Book, 2009). There is substantial evidence that the claiming of dependents is an element of tax compliance that creates perverse incentives for filers (LaLumia & Sallee, 2013; Liebman, 2000; McCubbin, 2000). A taxpayer is always better off if he or she can claim a child through the combined advantages of the dependent exemption, the head-of-household filing status, the EITC, and the Child Tax Credit (conditional on meeting other eligibility requirements) (Jones & O'Hara, 2016). Thus, taxpayers who wish to get the most out of the system may stretch or violate admittedly complex tax laws, and they may provide erroneous information to preparers (Masken, et al., 2009). In sum, the question of the “preparer effect” in tax fraud or incorrect credit payments is an open and a complicated one.

3. Data

3.1 Data description, sample description, and summary statistics

The data I use stem from a joint statistical contract between the U.S. Census Bureau and the IRS. 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. Using survey data, I am able to determine the members of the population who appear to be eligible, regardless of whether they file a Form 1040. The earliest year of data used to compute the take-up rate is 2005, and the process of take-up calculation is reported in detail for that year by Plueger (2009). The process has changed somewhat in subsequent years, mainly in the refinement of income measurement.

The tax data included in the project are, for each year, Form 1040 personal 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 records. From 2008 forward, but with the exception of tax year 2011, these records also include information on how a Form 1040 was filed (by a tax preparer, by the taxpayer online, or by the taxpayer via paper) and whether a Form 1040 filing included a refund anticipation product (RALs and RACs are coded separately).

The survey data used in this project are yearly Current Population Survey Annual Social and Economic Supplements (CPS ASEC) from 2009 to 2015, matched at the individual level for the corresponding tax year with the IRS data (that is, 2008 to 2014). Records are linked at the Census Bureau using a process whereby individuals in each data set were given a unique, protected identification key, called a 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 identifier is assigned based on SSN. Identifier placement is close to 100 percent in the case of administrative tax records with an SSN. For records without an SSN, personally identifiable information such as name, address, and date of birth is used in probabilistic matching against a reference file to assign PIKs. Personal information is then removed from each data set before they may be used for research purposes. For the EITC estimation project, I also remove persons whose income and wage values were imputed in the CPS ASEC, as initial EITC eligibility determination is dependent on these values.

Each CPS ASEC, administered annually in March, includes questions regarding family structure and earnings that can be used to estimate tax filings for the year preceding the survey. I rely on identifiers in the survey data to group people into tax filing units and to establish the relationship between everyone in the household and the household reference person (also known as the “householder”). The householder is presumed to be the primary tax filer for the identified tax unit; in the case of multifamily households, family identifiers are similarly used to distinguish separate units. In a later adjustment, the tax information on eligibility is transferred from the householder to the spouse if it was the spouse who filed. For each filing unit, I then assign secondary filer status and dependents. This strategy relies on an expanded relationship variable, which reports the exact relationship of a person to the householder (spouse, parent, child, sibling, etc.). Children of the householder are further identified using a parent pointer, which designates the parent’s identifier for each child in the household. Variables on tax unit earnings, income, and dependent support requirements (which, in the case of the EITC, is based on where a child lived for the tax year) are first taken from the survey data. Then, values from the tax data, when available, are swapped in for the survey values, and eligibility is refined based on these new values.

This strategy provides a denominator expressing everyone who is eligible for EITC in the tax year, regardless of whether they file a Form 1040. The numerator of the target take-up rate is the subset of eligible tax filers who actually file a Form 1040 and claim the credit. The incorrect payment rate is those tax filers who do not look eligible based on their modeled eligibility but who claim and are paid EITC in the tax year. The next subsection describes in detail the process of determining the incorrect payment rate and compares it to an IRS-calculated rate derived from audits. In what follows, I define the incorrect payment rate (described at length in section 3.2) as the number of CPS ASEC ineligible persons receiving EITC in a tax year divided by the number of taxpayers linked with the CPS ASEC.

The sample population is CPS ASEC respondents who file taxes. I observe EITC receipt, tax-preparation method, and refund anticipation product purchase for this population. Also available is a rich set of filer characteristics, including age, sex, race, Hispanic origin, foreign-born status, filing status (defined as single, head of household, or married[5]), the number of dependent children claimed on the Form 1040, educational attainment, adjusted gross income (AGI), investment income, self-employment status, urban zip code. I also include measures of two social programs that are correlated with EITC receipt: TANF and SNAP participation.

There is always a concern, when linking data sets, that a prevalence of mismatch leads to biased estimates. In the project described in this paper, the population of interest is 1040 filers. There is selection into this population; however, Form 1040 filing is a precondition for each of the main variables in the analysis (tax-preparer use, product purchase, and incorrect EITC payment). Therefore, any statements about the results of the analysis should be taken to apply only to tax filers. The analysis also relies on variables from survey data; the resulting linked files for each year do not constitute the population of 1040 filers or the original random sample of the survey. To address this issue, I calculate the probability that a CPS respondent received a PIK[6] and use the inverse of this probability to reweight the CPS ASEC person weight and replicate weights. Because Form 1040 observations receive PIKs close to 100 percent of the time, calculating the probability that a CPS ASEC respondent receives a PIK amounts to calculating the probability that a tax-filing CPS ASEC respondent is found in the Form 1040 data. The Appendix reports on how well this strategy covers the number of Form 1040 filers in each tax year. The weighted mean and standard error for all variables used in the analysis are shown by tax year in Table 1.

---

[5] Married filing separately is included in the "single" category. Married persons filing separately are not eligible for EITC.

[6] The placement of PIKs for the CPS ASEC in each year is around 90 percent.

The number of linked CPS ASEC persons over time declined over the period, both because the number of persons surveyed decreased and because there is more earnings non-response over time. When weighted, however, the number of linked CPS ASEC persons increased. There was a slight decrease in the rate of using a tax preparer, with about 58.8 percent using a preparer in 2008 versus 55.3 in 2014. Meanwhile, filing online with tax-preparation software increased from 32.0 percent to 40.6 percent. These numbers are in line with reports by the Director of IRS to Congress in April 2014, which reported 56 percent of returns were filed by a tax preparer and 34 percent were filed with tax preparation software.[7] The latter rate is higher in my data because I am including as "online" those who prepared their taxes using tax-preparation software, but then printed out the return and mailed it. Meanwhile, those writing out paper returns by hand and mailing them dropped from 9.3 percent to 4.2 percent. The mean of RAL use decreased from about 5.6 percent in 2008 to 0.1 percent by 2014, while RAC use increased from 8.1 percent to 13.0 percent. The use of any product decreased slightly: from 13.8 percent in 2008 to 13.0 percent in 2014.

Tax preparers include anyone who receives a preparer identifier from the IRS. Because there are no rules regarding licensing, a tax preparer may be a Certified Public Accountant, a volunteer preparer, an employee of a large tax preparation company, a used car salesperson, or a travel agent.[8] Except for 2014, I am only able to identify that a tax filing was made by a preparer, and not whether the preparer was volunteer. For 2014, however, the number of taxpayers who received volunteer assistance and were linked to the CPS ASEC was fewer than the IRS disclosure threshold, indicating that the incidence of volunteer preparation is extremely low.

---

[7] https://www.finance.senate.gov/imo/media/doc/Koskinen%20Testimony.pdf.
[8] Mother Jones, *Secrets of the Tax Prep Business,* http://www.motherjones.com/politics/2011/04/gary-rivlin-tax-prep-refund-anticipation-loan; e-mail correspondence with Dean Plueger, May 17, 2017

Figure 1 shows refund anticipation product use over time for the matched population. At the start of the period, a roughly equal number of taxpayers purchased a RAL through a preparer or a RAC through either a preparer or online. A very small number of taxpayers purchased an online RAL. Online RACs increased over the period, while there was a precipitous drop in preparer RALs in 2010 when the IRS eliminated the debt indicator. Preparers made up for the loss in RAL use by increasing RAC sales, which experienced a slower increase than the RAL decline between 2009 and 2010. By 2012, RALs purchased through either filing method decreased to zero (or numbers fewer than the IRS disclosure threshold, which are set to 0 in the graph). The rate of RACs purchased through a preparer did not differ significantly for the rest of the period, while those purchased online continued their upward trend.

3.2 Incorrect payment determination

Figure 2 graphically shows the EITC eligibility and credit structure as of 2014. A filer may be precluded from eligibility in several ways besides an absence of earnings. Earning investment income more than the allowed maximum precludes eligibility. A person with either earnings or income (AGI) beyond the highest threshold for their family structure is not eligible. If income is beyond the threshold, the person is ineligible even if earnings are within range and the person claims otherwise eligible children. This category of ineligible person would be those whose *combination* of children, filing status (single or joint), and maximum of earnings or income are outside of the eligible range.[9] Clearly, any person with earnings or income outside the maximum permitted in the tax year for any type of filer would be ineligible. Finally, persons without

[9] For example, a married filer with three children may have up to $52,427 in AGI compared with a single person, who must have below $46,997.

dependents are eligible for EITC only if they meet other eligibility requirements and are between the ages of 25 and 64.

A key question for the validity of the research I am presenting is whether I identify EITC recipients who are, in truth, ineligible. To be certain, I would have to require EITC recipients to undergo an audit. In the absence of an audit, I rely on comparisons between the tax data and survey data. Depending on the nature of the eligibility parameter in question, either the information reported on tax returns or the survey data is held to be the more likely measure.

The easiest determination of ineligibility occurs when there is an internal validity problem with the tax filing. For example, filers who report zero earned income on the 1040 (including Schedule C earned income or Schedule SE earned income) and have no W-2 earnings are ineligible, regardless of whether they meet other program parameters. A very few such filers claim EITC for the tax year and receive the credit.

The number of EITC eligible children claimed is a more difficult problem. The EITC conditions on child eligibility require that a qualifying child be related to the claimant biologically (son, daughter, grandchild, sibling, etc.) or be a formally adopted or foster child. A qualifying child must be either less than age 19 or less than age 24 and a full-time student, or permanently and totally disabled (at any age). Finally, the child must have lived with the claimant for more than half of the tax year. Thus the key determination in connecting potentially qualifying children to filers is that they live in the same household and should be present in the survey roster.

As outlined in Appendix B of Plueger (2009), the first step in determining which children qualify a filer for EITC is to create a qualifying-child unit identifier that collects every potentially qualifying child and every filer in the household based on survey responses. Then, the true number of children claimed for EITC by each filer on the Form 1040 is compared with the claiming that was previously determined for that filer using the survey variables. For example, say filer A and filer B live in the same household (and they do not file together). If we expect, based on the survey responses, that filer A should claim a certain number of children for EITC, but he claims fewer children on the Form 1040, we determine that there are "surplus" children in the household. If filer B then claims more children than he is expected to claim based on the survey responses, it seems reasonable that the "surplus" children have been appropriately claimed by filer B. When a filer claims more children than expected, I swap in the Form 1040 value for the survey value only when there are surplus children in the household. Otherwise, I assume that the claiming is not appropriate, and I take the previously determined survey value for the number of children claimed.

Income also undergoes adjustment based on a comparison of survey and tax data. For earnings, I swap in W-2 values for survey values when W-2 information is available for the filer. If it is not available, I swap in Form 1040 wage and salary information. W-2 information on earnings is preferred over Form 1040 information because of the well-documented problem of earnings overreporting for the EITC, mainly through self-employment earnings reports (Saez, 2010).

Finally, because filers often underreport investment income (Johns & Slemrod, 2010), I take the larger of survey-reported or Form 1040-reported investment income. Of those considered ineligible and incorrectly paid EITC due to investment income, nearly all reported investment income to the IRS that was less than their survey reports. The remainder actually claimed investment income on the Form 1040 that made them ineligible for EITC.[10]

[10] The actual number who claimed investment income more than the threshold on the Form 1040 was smaller than the disclosure threshold.

Because the population of interest for this paper is tax filers, these choices cover all of the observations in the analysis data. An instance of incorrect payment occurs when a filer receives a larger EITC credit than the estimated eligible value, with differences of $100 or less ignored. When population weights are used (reweighted to account for the probability of PIK placement), the resulting number of incorrectly paid EITC recipients in each year is close to that reported through the IRS audit process, with an incorrect payment rate for all EITC receivers of 22 to 25 percent.[11] In the Appendix, I report on categories of ineligibility and how these categories, incorrect EITC payment, and preparer and product use relate to one another.

Incorrect payment rates are intensified by the use of refund anticipation products. Figure 3 shows how the combination of tax preparation and product is associated with incorrect payment. To avoid disclosure issues, I have combined the counts of RALs and RACs into a single product, and graphed incorrect payment for five groups defined by whether they used a tax preparer, filed online, or filed by paper, and whether they purchased a product. Those filing via paper, by definition, do not purchase a product.

The graph shows that incorrect payment rates are largely associated with refund anticipation product use. Using a preparer and purchasing a product is associated with the highest incorrect payment rate (around 12 percent in most years, but 16 percent in 2010). Online users purchasing a product have an incorrect payment rate of about 8 percent, while both types of filers have incorrect payment rates of about 3 percent when no product is involved.

Incorrect payment and product use is likely codetermined by low income; the different levels of incorrect payment we see in Figure 4 may be driven by low-income taxpayers who turn to a combination of incorrect payment and product because they are liquidity constrained. Table

[11] https://www.irs.gov/pub/irs-soi/EITCComplianceStudyTY2006-2008.pdf. Here, the denominator is all tax filers rather than all EITC recipients, for an incorrect payment rate for all filers of 4.4 percent.

4 reports incorrect payment rates for each filing type, broken out by income categories. Over the period, those receiving a product from a preparer had the highest rates of incorrect payment regardless of income. Meanwhile, those receiving a product online showed lower levels of incorrect payment at each level of income. Those not receiving a product but using a preparer had incorrect payment rates between 1.6 and 6.2 percent, compared with 8.3 and 17.4 percent for the group buying a product.

The preceding look at the data provides some strong evidence that product use combined with paid preparer use has an association with incorrect EITC payment. However, the story so far cannot account for selection—perhaps taxpayers seek out paid preparation with the express purpose of claiming incorrectly and receiving quick refund money. The next section discusses an econometric model that will address the issue of selection.

4. Econometric model

As outlined in the preceding sections, tax preparers clearly have an incentive to sell refund anticipation products, and they receive a higher price when the refund is larger; meanwhile, many taxpayers, based on their refund status, financial situation, and liquidity constraints, have dual incentives to buy them and to claim large refunds. Both the tax preparer's and taxpayer's incentives may become perverse vis-à-vis tax evasion or avoidance. If we see that incorrect payment occurs more often in the presence of a product and when a tax preparer is involved, how do we separate out demand effects from supply effects? Taxpayers with intent to defraud may choose to use a tax preparer in the hopes of avoiding an audit. Meanwhile, employees of tax preparation services are pushed to sell refund anticipation products, and their bonus compensation is dependent upon these sales.[12] In the absence of some exogenous change to either supply or demand, it is impossible to disentangle these effects.

When the IRS announced that it would no longer provide the debt indicator for refund anticipation loans, it was August 2010 and applied to the upcoming 2011 tax-filing season. In other words, the debt indicator removal was an unexpected move that created a price shock for the tax-preparation industry. Suddenly, this type of loan was now unsecured, which increased the marginal cost of the product in relation to its price. Banks would no longer support the loans, which meant that the tax-preparation industry withdrew quickly from the product. The industry's response to this shock can be clearly seen in Figure 1—an abrupt departure from the RAL product offering to the lower-cost RAC in 2010.

Essentially, preparers were compelled to sell RACs in 2010 to customers who previously would have purchased RALs, as shown in Figure 2. As discussed in section 2, RACs vary in price dependent on their "add-in" fees, and the highest-priced products are sold to filers with large refunds that depend on the EITC and other credits. Given little time to change the baseline price of the RAC, individual tax preparers could instead make up some of the price difference through add-on fees. To do so, preparers would have to induce a higher proportion of RAC purchasers to file for EITC. The removal of the debt indicator thus provides an opportunity to estimate a "tax preparer" effect in the incorrect payment of EITC, separate from the filer effect. In an ideal world, I would have data that would allow me to calculate the price differences

[12] www.hsgac.senate.gov/download/ws-pat-eckelberry.pdf: "While I was at H&R Block, management pushed very hard to require preparers to use the client - preparer trust relationship to sell such products as Refund Anticipation Loans (RALs), Refund Anticipation Checks (RACs), and Peace of Mind (POM) guarantees . . . Tax preparers received 15 percent of the system administration fee that was paid to H&R Block for completing the bank-related paperwork for a RAL. In addition, the number of RALs, RACs etc. sold was a part of our performance reviews. I have also provided a copy of my performance evaluation to the Subcommittee, which included a category for financial products such as RALs and RACs."

between the two products from year to year. In the absence of actual prices, I use the year of the shock to detect differences in product use and incorrect payment.

I examine the impact of the removal of the debt indicator on the incorrect payment rate using a model in which I have two treatment groups (those buying a product when filing online or through a preparer); two corresponding control groups (those not buying a product and filing online or through a preparer); and a final baseline group consisting of those who file by paper (these never receive a "treatment," since "product"=0 for every member of this group). In examining the differential behavior of the group using a preparer and buying a product in 2010, I can estimate a preparer effect on the improper payment rate.

The choice of online filers as a comparison group makes the assumption that these filers are a comparable control for those who use a tax preparer with respect to the price shock in the RAL market. If the price shock induced a change in the supply of filers to online preparation, this use of the online group as the control would be suspect. However, Figure 2 provides evidence that there was no spike in online filing/product purchase in 2010. There is a steady year-by-year increase in online RAC purchases from 2009 to 2010, but the 2010 number is not statistically different from 2009. Evidence indicates that the price shock had little influence on the demand of products—filers used tax preparation and products at a similar rate but had no choice other than to accept the RAC over the RAL due to the preparers' supply decision. "Product," then, can cover either a RAL or a RAC, in that from the supplier's point of view, they differ only in their marginal cost and price structure.

The specification is as follows:

$$y_{its} = \alpha + \beta_1 prepXproductXyear_{2009} + \cdots + \beta_5 prepXproductXyear_{2014} + \gamma_1 onlineXproductXyear_{2009} + \cdots + \gamma_5 onlineXproductXyear_{2014} + \delta prepXproduct + \varphi onlineXproduct + \theta prep + \rho online + \tau_1 year_{2009} + \ldots + \tau_5 year_{2014} + \sigma_s + X'_{its}\beta + \epsilon_{its}$$

In this linear probability model, the dependent variable is the probability of incorrect payment of EITC. In a second specification, I also examine the value of EITC overpayment. Each year fixed effect captures changes for the baseline group over time: paper filers without a product. *Prep* and *online* capture the difference in the incorrect payment rate for these groups in the absence of a product in the base year, while the interactions of these two states with *product* capture the difference in incorrect payment for these groups in 2008 when a product is used. Finally, the triple interaction with the two filing groups, product indicator, and year dummies gives the year-by-year difference in incorrect payment by group and product use. The coefficients of interest are β_2 and γ_2, the triple-interaction coefficients for 2010. The difference between them captures the impact of the debt indicator removal, which I posit is a "preparer effect," on the incorrect payment rate. X includes the covariates reported in Table 1, and σ are state fixed effects, which are included to control for differences in state EITC, state-level lending and tax preparer regulations, and state-specific labor-market conditions. Standard errors are clustered at the state level, and the models are weighted using the recalculated CPS person weights.

5. Results

Table 3 shows the key results of the analysis. For simplicity, the table focuses on the impact of the 2010 removal of the debt indicator, reporting only the direct and interacted effects for 2010, the preparer and online dummies, and the triple interaction of these dummy variables with the product indicator.[13] Looking at model 2, which includes covariates: compared with paper filers, filers who used a preparer were about 1.1 percentage points more likely to be incorrectly paid EITC in 2008, and online filers were 1.0 percentage points more likely. The inclusion of a product in 2008 adds another 13.4 percentage points if the filer used a preparer, and another 5.1 percentage points if filing online. There was an increase in 2010 for those filing online and a decrease for those filing with a preparer, but the decrease for preparer filings is not statistically different from zero. There is little change between estimates from model 1 to model 2—the estimates withstand the inclusion of demographic covariates, although the coefficients on the interaction of filing type and product are slightly smaller.

[Table 3 about here]

The hypothesized "preparer effect" in the interplay of refund anticipation products and incorrect payment is the difference between the triple interactions for each model. Those filing with a preparer and purchasing a product saw an additional 6.2 percentage points in the improper payment rate in 2010; meanwhile, those filing online and purchasing a product in 2010 experienced a 3.0 percentage point increase. An F test that the coefficients are equal rejects the null at the 10 percent level. While the full model gives a better demonstration of what is going on with tax preparation and product use for all filers, we can also restrict the analysis sample to only those who file online or through a preparer. Model 3 shows the results of this exercise, which turns the previous model into a standard triple-difference. In this case, I can include the product dummy and its interaction with each year indicator, which will capture the effect for the base group: "online." Here we can see that, compared with online filers, those filing with a preparer in

[13] Appendix table A4 and A5 report the full sets of results from this section, with the exception of the state fixed effects.

2008 had an incorrect payment rate of less than a percentage point higher, but when a product is added, the increase in rate is 4.7 percentage points. Add on the effect of filing by preparer in 2010 and purchasing a product, and the increase is an additional 3.3 percentage points, which is approximately the difference between the coefficients from the previous model 2. This estimate is again marginally statistically significant, with a *p* value of about 0.06. In this model, filers who use a preparer but do not purchase a product saw a decrease over 2008 in the incorrect payment rate, perhaps indicating that the incentive to file incorrectly decreased for clients who did not purchase a product in 2010.

Finally, I look at a low-income subset of the sample. Although I control for log income and income-squared in model 2,[14] it should be of interest to see whether effects are concentrated at the range of income where filers may be eligible for EITC along the income parameters. In contrast with expectations, the coefficients of interest on the two interaction terms are smaller and both are not statistically different from zero. An F test does not reject the null that they are the same. There does seem to be a higher likelihood of incorrect payment for this subset overall compared with the full sample, but the 2010 removal of the debt indicator did not appear to affect behavior in this subset more strongly than in the full sample.

Thus far, the modeled decision is binary: either a filer incorrectly claims EITC or does not. The next set of results examines the value of EITC overpayment and whether this varies by filing type, product purchase, preparer use, and year. Table 4 provides the mean and standard error, by year, for overpayment using all filers in the sample. Note that the largest possible overpayment value is \$6,143—the maximum credit value in 2014. The dollar values are nominal (further analysis employs year fixed effects). The difference in means between online and preparer filings when a product is included is statistically significant, with preparer-plus-product

[14] This choice reflects the fact that over the range of the credit, value increases and then decreases in income.

filings associated with overpayments of $250 to $390 in each year (2010 is the year associated with the largest difference). Meanwhile, difference in overpayments between online and preparer filings are a few dollars and not statistically different.

Examining this difference in the same econometric model used for estimating the probability of incorrect payment presents some challenges. Considering that incorrect payment occurs only 10 percent of the time over all taxpayers, the dependent variable in question—"overpayment"—is heavily populated with zeros. This requires a careful consideration of the choice of model. After investigation, a generalized linear model with a log link provided the most stable results and best fit. Such models are supported for the context of dependent variables with zero inflation (Nichols, 2010), especially when the zeros in question are true values and not the result of censoring. Table 5 reports the results from this analysis, using a similar difference-in-differences approach and the same set of covariates. The dependent variable is the dollar amount of overpayment, where those who were paid within $100 of their eligible amount were coded as zero. Because the value of the EITC is mechanically dependent on income, looking separately at the low-income sample is less meaningful in this case, and the model results are not reported for the low-income sample.

The results reported in the table are coefficients; exponentiated, the coefficient on "preparer" indicates that preparer and online filings are associated with an approximate 2 percent increase in the overpayment value compared with paper filings (when all other variables are evaluated at the mean). Moreover, there is an added positive effect on overpayment for either type of filer upon product use.

For models 1 and 2, neither triple-interaction term is statistically significant. For model 3, however, the difference between preparer filings with a product and online filings with a product in 2010 is statistically meaningful. The interaction terms reported in Table 5 cannot be interpreted directly (Ai & Norton, 2003). To evaluate the marginal effect of "product" on the value of incorrect payment, I calculate each year's marginal effect (in dollar rather than log terms) for product separately for prepared filings and online filing, using the results of model 3 (again, all other covariates are held at their mean values). These effects are graphed in Figure 4. There is a spike in 2010 in the strength of the marginal effect for prepared filings compared with online filings, with a difference of about $125 between the two. The changes in marginal effect for prepared and online filings are statistically different from one another at the 5 percent confidence level when comparing 2008 to 2010 and 2009 to 2010 (Chi-squared statistics of 7.19 and 4.72, respectively). On the downward slope, the difference is statistically significant for the difference in changes between 2010 and 2012 (Chi-squared of 5.21). Marginal effects for both categories of filer increase over time from 2012 onward, but the differences in marginal effects are not statistically significant between 2012 and 2014.

Together, the two sets of results from Tables 3 and 5 confirm the involvement of paid preparers in the incorrect claiming of EITC, with higher overpayment opportunities inducing a greater propensity to incorrectly claim in 2010 for preparer plus product filings. The availability of refund anticipation products appears to induce a response in the tax filer, whether filing online or using a preparer, but the additional propensity and value seen in the triple interaction for 2010 supports the story that preparers respond to the incentive separately.

6. Specification checks

There are two issues that warrant investigation when it comes to the results just presented. The first is the concern that customers of preparers differed in their preference for a RAL over a RAC in relation to their desire to incorrectly claim and the value of the incorrect payment. In other words, because the preceding results depend on the treatment of RALs and RACs being such close substitutes that customers of preparers were willing to accept a RAC instead of a RAL, it is important to demonstrate that RAL and RAC customers in the pre-period did not select into the product type based on an intention to incorrectly claim. Because preparers withdrew from the RAL market, RAL versus RAC use cannot be explored as separate behavior after 2010.

To assess this issue, I examined tax filers' selection into a RAL or a RAC before 2010 based on whether the filer incorrectly claimed and the value of their EITC. The results are reported in Table 6. The sample is those filers who used a tax preparer before 2010. The dependent variable is a 1 if a filer chose a RAL, and a 0 if the filer chose a RAC. Incorrect payment is defined as before, and it is interacted with the value of the EITC, which in this case is logged. A dummy variable is included that equals 1 if a filer received no EITC (mean logged EITC replaces the continuous variable for these observations). Covariates are the same as those used in the models reported previously, and standard errors are clustered at the state level. Each of the main variables in this examination were not associated in a statistically significant way with selection into a RAL versus a RAC. All indications are that customers' selection into a RAL or a RAC did not vary in terms of incorrect claiming and the value of the incorrect claim before 2010.

A second concern is whether incorrect payment of EITC was due to volume, with tax preparers handling more EITC claims that included a product in 2010 than were made online or by paper. This is examined in Table 7, which reports the same Model 2 as in the main results, but with a dependent variable coded 1 if a tax filer received EITC (regardless of eligibility).

Overall volume of EITC claimants who bought a product through a preparer did not experience a statistically significant change in 2010. While there are more EITC claimants who file via preparer or online versus paper, and buying a product is associated with receiving EITC, there was no increase in traffic in EITC claims for the effect of interest: Both triple-interaction terms are not statistically different from zero. The results from tables 6 and 7 indicate that the greater propensity for preparers to file incorrectly in 2010, and for credits of greater value, are not due to filers' selection into a particular kind of product or for changes in filer type specific to 2010.

7. Conclusion

Each year, taxpayers rely on the refund money due to them from the tax and transfer system; receipt of refunds is especially important to filers who receive the EITC. Since the advent of e-filing, tax preparers have been absorbing a proportion of this refund money. Some see tax preparers as providing a valuable service, in that they can guide taxpayers through tax complexity, making sure they take advantage of the credits and exemptions they are eligible for. Yet, through high fees for preparation and—especially—for refund products, tax preparers make a substantial profit on low-income taxpayers.

These profits create perverse incentives for tax preparers to move filers into the category "eligible" when it comes to EITC. I show that rates of incorrect EITC payment and EITC overpayment values are higher for preparer-made filings compared with online filings when a refund anticipation product is purchased; however, these differential rates may be accounted for due to choices made by the filer. Using an exogenous price shock to a particular refund anticipation product, I estimate a "preparer effect" in the incorrect payment of EITC: in response to this shock, there was an increase in incorrect payment for filers who used a preparer and bought a product in the year of the shock. Moreover, the value of EITC overpayment was higher for preparer-plus-product filings in this year. This "preparer effect" was statistically different from the effect calculated for online filers in the same year who bought a product—marginally so in the case of any incorrect filing and at conventional levels for overpayment.

To my knowledge, this is the first research able to directly estimate a tax-preparer effect on tax evasion (a broader issue certainly not confined to EITC recipients). The results of this paper should help inform ongoing policy discussions on how to address the incorrect payment rate of EITC while simultaneously protecting eligible receivers, as well as discussions regarding tax preparer regulation and licensing.

References

Ai, C. & Norton, C., 2003. Interaction terms in logit and probit models. *Economics letters,* 80(1), pp. 123-129.

Barr, M. S. & Dokko, J. K., 2008. Third-party tax administration: The case of low-and moderate-income households.. *Journal of Empirical Legal Studies,* 5(4), p. 963–981.

Berube, A., Forman, B., Burns, M. & Kim, A., 2002. *The price of paying taxes: How tax preparation and refund loan fees erode the benefits of the EITC.,* s.l.: s.n.

Bitler, M., Hoynes, H. & Kuka, E., 2014. *Do in-work tax credits serve as a safety net?,* s.l.: NBER Working Paper.

Book, L., 2009. Refund Anticipation Loans and the Tax Gap. *Stanford Law and Policy Review,* 20(85).

Davis, S. L., 1998. *Unbridled Power: Inside the Secret Culture of the IRS.* New York: HarperBusiness (HarperCollins).

Dewees, S. & Parrish, L., 2009. *Borrowed Time: Use of Refund Anticipation Loans Among EITC Filers in Native American Communities.,* s.l.: First Nations Development Institute, Center for Responsible Lending, and First Nations Development Institute. Native Assets Research Center..

Eissa, N. & Hoynes, H., 2008. *Redistribution and Tax Expenditures: The Earned Income Tax Credit,* s.l.: s.n.

Elliehausen, G., 2005. *Consumer use of tax refund anticipation loans.,* s.l.: Credit Research Center Monograph.

Feltner, T., 2007. *Debt Detour: The Automobile Title Lending Industry in Illinois.,* Chicago, IL: Woodstock Institute and the Public Action Foundation.

Finklestein, A., 2007. *E-ztax: Tax salience and tax rates,* s.l.: NBER Working Paper.

Holt, L., 2009. *Testimony by Leonard Holt, CHH Small Firms Services, at the IRS Return Preparer Review Forum.* s.l.:s.n.

Holt, S., 2011. *Ten years of the EITC movement: Making work pay then and now.,* s.l.: s.n.

Johns, A. & Slemrod, J., 2010. The Distribution of Income Tax Noncompliance. *National Tax Journal,* 63(3), p. 397.

Jones, M. R., 2016. *A Loan by Any Other Name: How State Policies Changed Advanced Tax Refund Payments,* s.l.: s.n.

Jones, M. R. & O'Hara, A., 2016. Do Doubled-up Families Minimize Household-level Tax Burden?. *National Tax Journal,* 69(3).

Kaplow, L., 1996. How Tax Complexity and Enforcement Affect the Equity and Efficiency of the Income Tax. *National Tax Journal,* 49(1), p. 135.

Kopczuk, W. & Pop-Eleches, C., 2007. Electronic Filing, Tax Preparers and Participation in the Earned Income Tax Credit. *Journal of Public Economics,* 91(7-8), pp. 1351-67.

LaLumia, S. & Sallee, J. M., 2013. The Value of Honesty: Empirical Estimates from. *International Tax and Public Finance,* 20(2), p. 192–224.

Levy, A. H., 2015. Believing in Life After Loving: IRS Regulation of Tax Preparers.. *Florida Tax Review,* Volume 17, pp. 437-669.

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

Masken, K., Mazur, M., Meikle, J. & Nord, R., 2009. *Do Products Offering Expedited Refunds Increase Income Tax Non-Compliance?.* s.l., s.n.

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

Melzer, B. T., 2011. The Real Costs of Credit Access: Evidence from the Payday Lending Market. *The Quarterly Journal of Economics,* 126(1), pp. 517-555.

Nichols, A., 2010. Regression for nonnegative skewed dependent variables. *BOS10 Stata Conference,* Volume 2.

Nichols, A. & Rothstein, J., 2015. The Earned Income Tax Credit. In: *Economics of Means-Tested Programs in the United States, Vol. 1.* s.l.:University of Chicago Press, pp. 137-218.

Plueger, D., 2009. *Earned Income Tax Credit Participation Rate for Tax Year 2005,* s.l.: s.n.

Rothstein, D. & Black, R., 2015. *Improving the Tax Preparation Experience,* New York: s.n.

Saez, E., 2010. Do Taxpayers Bunch at Kink Points?. *American Economic Journal: Economic Policy,* 2(3), pp. 180-212.

Theodos, B. et al., 2011. *Who Needs Credit at Tax Time and Why: A Look at Refund Anticipation Loans and Refund Anticipation Checks,* s.l.: Urban Institute.

Wu, C. C. & Best, M., 2015. Taxpayer Beware: Unregulated Tax Preparers and Tax-Time Financial Products Put Taxpayers at Risk.. In: *NCLC/CFA 2014 Refund Anticipation Loan Report.* Boston: National Consumer Law Center.

Wu, C. C., Del Rio, D., Iwanisziw, A. & Skillern, P., 2011. *Tax Time 2011: Shopper Testing In New York and North Carolina Finds Continuing Problems With Tax Preparers,* Boston: National Consumer Law Center.

Wu, C. C. & Feltner, T., 2014. It's a Wild World: Consumers at Risk from Tax-Time Financial Products and Unregulated Preparers.. In: *NCLC/CFA 2013 Refund Anticipation Loan Report..* Boston: National Consumer Law Center.

Zinman, J., 2010. Restricting Consumer Credit Access: Household Survey Evidence on Effects Around the Oregon Rate Cap. *Journal of Banking & Finance,* 34(3), pp. 546-556.

Tables and Figures

Table 1. Summary statistics for analysis variables by tax year, weighted

	2008		2009		2010		2012		2013		2014	
Variable	mean	se	mean	se	mean	se	mean	se	mean	se	mean	se
Preparer	0.588	0.003	0.584	0.003	0.582	0.003	0.569	0.003	0.562	0.003	0.553	0.003
Online	0.320	0.002	0.336	0.003	0.358	0.003	0.380	0.002	0.391	0.003	0.406	0.003
Paper	0.093	0.002	0.079	0.001	0.060	0.001	0.051	0.001	0.047	0.001	0.042	0.001
Any product	0.138	0.002	0.139	0.002	0.138	0.002	0.142	0.002	0.127	0.002	0.130	0.002
RAL	0.056	0.001	0.048	0.001	0.013	0.001	0.001	0.000	0.000	0.000	0.001	0.000
RAC	0.081	0.001	0.091	0.002	0.125	0.002	0.141	0.002	0.127	0.002	0.130	0.002
Paid EITC	0.177	0.002	0.191	0.002	0.193	0.002	0.192	0.002	0.191	0.002	0.187	0.002
Age	44.493	0.071	45.078	0.066	45.461	0.069	45.585	0.066	45.725	0.073	45.886	0.071
Sex (Male=1)	0.651	0.002	0.651	0.002	0.648	0.002	0.637	0.002	0.635	0.002	0.634	0.002
Single filer	0.234	0.002	0.235	0.002	0.236	0.002	0.234	0.002	0.233	0.002	0.238	0.002
Head of household	0.078	0.001	0.080	0.001	0.079	0.001	0.082	0.001	0.079	0.001	0.078	0.001
Married	0.439	0.003	0.437	0.002	0.433	0.003	0.429	0.003	0.431	0.003	0.426	0.003
Number of children	0.527	0.004	0.548	0.005	0.539	0.004	0.542	0.004	0.536	0.004	0.525	0.005
White	0.819	0.002	0.818	0.002	0.814	0.002	0.801	0.002	0.802	0.002	0.798	0.002
Black	0.114	0.001	0.115	0.001	0.118	0.001	0.120	0.001	0.117	0.001	0.118	0.001
Asian	0.042	0.001	0.044	0.001	0.044	0.001	0.050	0.001	0.053	0.001	0.055	0.001
Other	0.024	0.001	0.023	0.001	0.023	0.001	0.029	0.001	0.028	0.001	0.029	0.001
Hispanic	0.135	0.002	0.134	0.001	0.134	0.002	0.145	0.002	0.148	0.002	0.151	0.002
Less than HS	0.116	0.002	0.111	0.002	0.106	0.002	0.099	0.002	0.095	0.002	0.096	0.002
HS degree	0.283	0.003	0.289	0.003	0.282	0.002	0.273	0.003	0.267	0.003	0.266	0.002
Some college	0.300	0.002	0.295	0.002	0.299	0.002	0.307	0.002	0.304	0.003	0.302	0.003
BA/BS +	0.301	0.003	0.305	0.003	0.314	0.003	0.322	0.003	0.334	0.003	0.336	0.003
No AGI	0.006	0.000	0.004	0.000	0.007	0.000	0.005	0.000	0.004	0.000	0.005	0.000
Log AGI	10.357	0.009	10.348	0.007	10.399	0.007	10.397	0.007	10.438	0.007	10.459	0.008
Log AGI squared	109.577	0.134	109.198	0.128	110.459	0.129	110.254	0.133	111.104	0.132	111.714	0.147
No TANF	0.993	0.000	0.993	0.000	0.992	0.000	0.993	0.000	0.994	0.000	0.993	0.000
Log TANF value	0.048	0.003	0.049	0.003	0.057	0.003	0.051	0.003	0.047	0.003	0.055	0.004
No SNAP	0.939	0.001	0.923	0.002	0.910	0.002	0.907	0.002	0.903	0.002	0.904	0.002
Log SNAP value	2.616	0.007	2.714	0.008	2.788	0.009	2.806	0.011	2.824	0.011	2.809	0.010
Observations	59,741		58,568		56,284		55,275		51,401		48,227	
Weighted observations	130,203,505		127,939,439		127,430,267		131,635,761		130,800,083		131,308,372	

Source: Linked CPS ASEC-Form 1040 data, 2008-2010, 2012-2014

Note: 0.000 indicates that the values rounded to 0 but are not exactly 0.

Table 2. Rates of incorrect payment by income and filing type, weighted, all years

	Preparer with product	Online with product	Preparer, no product	Online, no product	Paper
Less than $10K	0.083	0.049	0.025	0.024	0.012
$10K to < $20K	0.125	0.091	0.046	0.055	0.038
$20K to < $30K	0.174	0.118	0.062	0.068	0.026
$30K to < $40K	0.156	0.088	0.051	0.054	0.022
$40K to < $50K	0.154	0.086	0.048	0.048	0.017
>=$50K	0.150	0.060	0.016	0.020	0.008

Source: Linked CPS ASEC-Form 1040 data, 2008–2010, 2012–2014. Cells show the incorrect payment rate by type of filer/product use and bins of income.

Table 3. Results of difference analysis, weighted estimates: dependent variable is the probability of incorrect payment

	(1) Baseline	(2) With covariates	(3) Online as comparison	(4) Low income
Preparer	0.020***	0.011***	0.001	0.011***
	(0.003)	(0.002)	(0.003)	(0.003)
Online	0.018***	0.010***		0.012**
	(0.003)	(0.003)		(0.004)
Preparer X Product	0.209***	0.134***	0.083***	0.129***
	(0.013)	(0.012)	(0.012)	(0.012)
Online X Product	0.078***	0.051***		0.056***
	(0.012)	(0.011)		(0.014)
Preparer X Product X 2010	0.060***	0.062***	0.033	0.061***
	(0.009)	(0.009)	(0.017)	(0.010)
Online X Product X 2010	0.029*	0.030*		0.030
	(0.013)	(0.013)		(0.016)
Preparer X 2010	0.003	0.001	-0.019***	0.004
	(0.003)	(0.004)	(0.004)	(0.006)
Online X 2010	0.023***	0.020***		0.029***
	(0.004)	(0.005)		(0.007)
Year = 2010	0.008*	0.010*	0.030***	0.012
	(0.003)	(0.004)	(0.004)	(0.006)
Product			0.050***	
			(0.011)	
Product X 2010			0.030*	
			(0.013)	
State FE	yes	yes	yes	yes
Income <$60,000	no	no	no	yes
Demographic covariates	no	yes	yes	yes
Test of $\beta_2 = \gamma_2$	3.23	3.61		2.30
Prob > F	0.079	0.063		0.136
Obs.	336,166	336,166	315,041	213,197

* p<0.05, ** p<0.01, *** p<0.001. Source: Linked CPS ASEC-Form 1040 data, 2008–2010, 2012–2014. Results from a difference-in-differences model comparing five groups: paper filers as the base category; preparer with a product; online with a product; and preparer and online with no product. Full results reported in Appendix Table A4. Standard errors clustered at the state level shown in parentheses.

Table 4. Weighted means and standard error of overpayment by year, product use, and filer type

	Product				No Product			
	Online		Preparer		Online		Preparer	
	mean	se	mean	se	mean	se	mean	se
2008	232.68	21.41	505.07	16.50	69.42	3.65	70.03	2.84
2009	247.01	17.63	565.89	17.54	83.56	4.32	81.72	3.24
2010	305.38	19.19	694.75	22.97	116.79	4.81	91.12	3.60
2012	297.19	20.20	536.69	19.64	102.94	4.46	80.76	3.17
2013	343.96	23.77	603.43	24.31	113.57	4.92	96.09	3.88
2014	320.57	22.15	659.36	29.68	109.56	4.87	97.31	4.01

Source: Linked CPS ASEC-Form 1040 data, 2008–2010, 2012–2014.

Table 5. Results of difference analysis, GLM with a log link, weighted estimates: dependent variable is the probability of incorrect payment

	(1) Baseline	(2) With covariates	(3) Online as comparison
Preparer	0.836***	0.663***	0.027
	(0.122)	(0.126)	(0.072)
Online	0.823***	0.641***	
	(0.131)	(0.131)	
Preparer X Product	1.861***	0.885***	0.171
	(0.082)	(0.055)	(0.113)
Online X Product	1.173***	0.714***	
	(0.128)	(0.122)	
Preparer X Product X 2010	0.050	0.117	0.323*
	(0.079)	(0.072)	(0.157)
Online X Product X 2010	-0.228	-0.205	
	(0.154)	(0.138)	
Preparer X 2010	0.100	0.012	-0.247*
	(0.166)	(0.185)	(0.097)
Online X 2010	0.351	0.258	
	(0.184)	(0.204)	
Year = 2010	0.159	0.232	0.492***
	(0.174)	(0.190)	(0.074)
Product			0.719***
			(0.122)
Product X 2010			-0.207
			(0.138)
State FE	yes	yes	yes
Demographic covariates	no	yes	yes
Obs.	336,166	336,166	315,041

* p<0.05, ** p<0.01, *** p<0.001. Source: Linked CPS ASEC-Form 1040 data, 2008–2010, 2012–2014. Results from a difference-in-differences model calculated using a GLM with a log link comparing five groups: paper filers as the base category; preparer with a product; online with a product; and preparer and online with no product. Full results reported in Appendix Table A5. Standard errors clustered at the state level shown in parentheses.

Table 6. Results of specification check: probability of purchasing a RAL versus a RAC before 2010

Variable	Incorrect payment	EITC amount (log)	Incorrect payment*EITC amount	No EITC
Coeff.	-0.075	0.015	0.011	0.051
SE	(0.089)	(0.008)	(0.011)	(0.060)

* p<0.05, ** p<0.01, *** p<0.001. Source: Linked CPS ASEC-Form 1040 data, 2008–2009. Results from model comparing filers who use a preparer and buying a RAL or a RAC. Number of observations: 17,363. Clustered standard errors shown in parentheses.

Table 7. Results of difference-in-differences analysis, weighted estimates: dependent variable = paid EITC

	(2) With covariates
Preparer	0.033***
	(0.005)
Online	0.032***
	(0.004)
Preparer X Product	0.206***
	(0.009)
Online X Product	0.057***
	(0.009)
Preparer X Product X 2010	0.005
	(0.008)
Online X Product X 2010	0.020
	(0.013)
Preparer X 2010	0.001
	(0.008)
Online X 2010	0.013
	(0.007)
Year = 2010	0.008
	(0.007)
State FE	yes
Demographic covariates	yes
Obs.	336,166

* p<0.05, ** p<0.01, *** p<0.001. Source: Linked CPS ASEC-Form 1040 data, 2008–2010, 2012–2014. Results from a difference-in-differences model comparing five groups: paper filers as the base category; prep with a product; online with a product; and prep and online with no product. Standard errors clustered at the state level shown in parentheses.

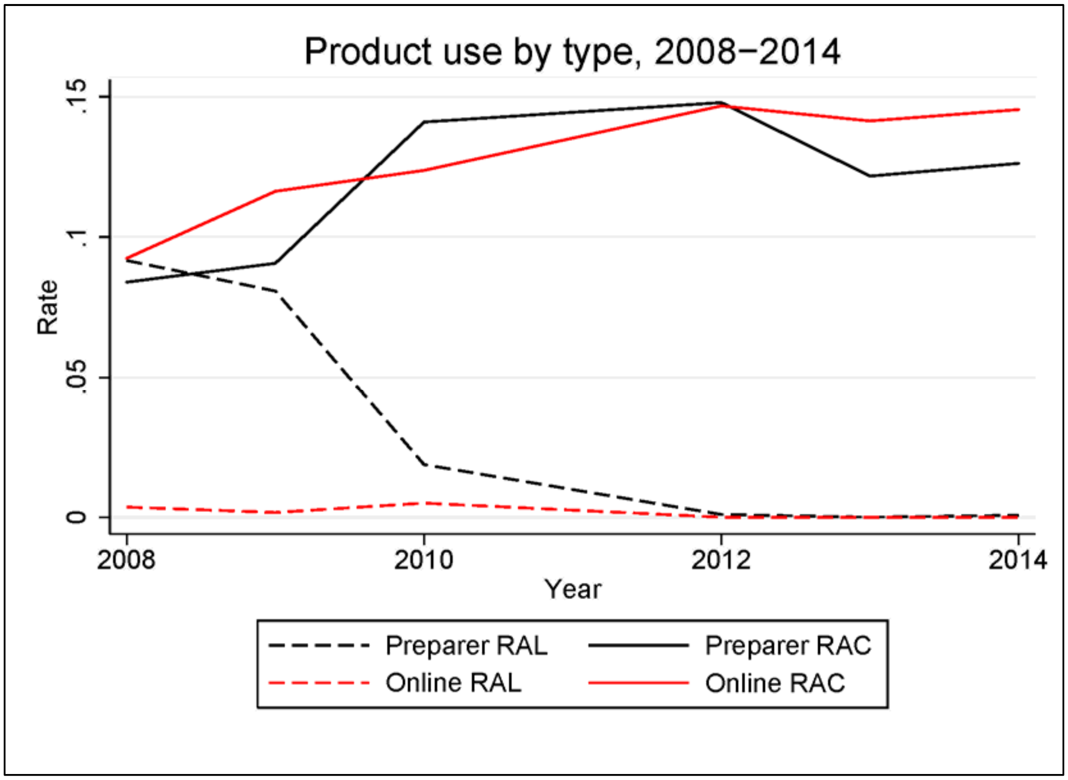

Figure 1. Source: Linked CPS ASEC-Form 1040 data, 2008-2010, 2012-2014, with year 2011 linearly interpolated.

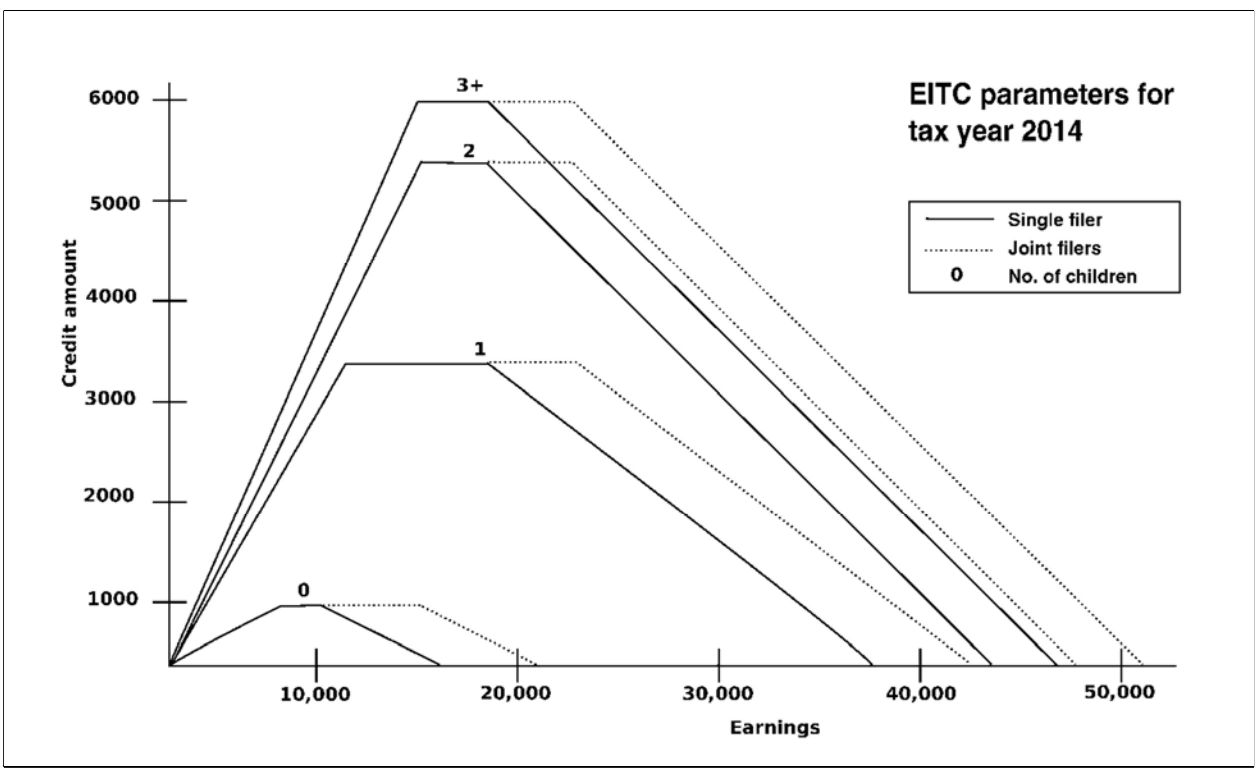

Figure 2. The EITC schedule for 2014. Source: Author's illustration created using program parameters outlined at irs.gov. Single filers include those who file as head of household.

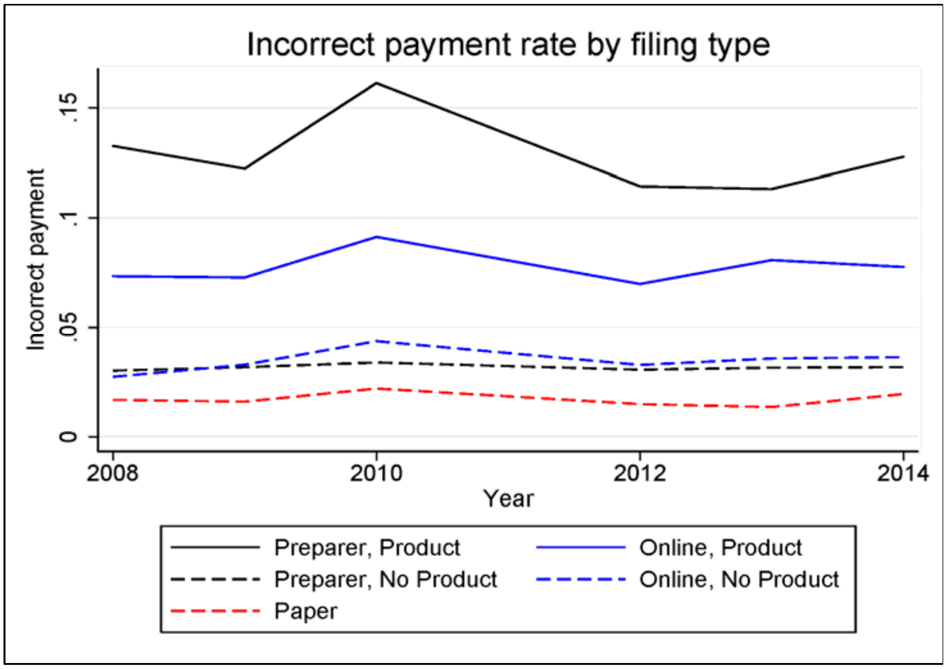

Figure 3. Incorrect payment rate by filing type. Source: Linked CPS ASEC-Form 1040 data, 2008–2010, 2012–2014, with year 2011 linearly interpolated.

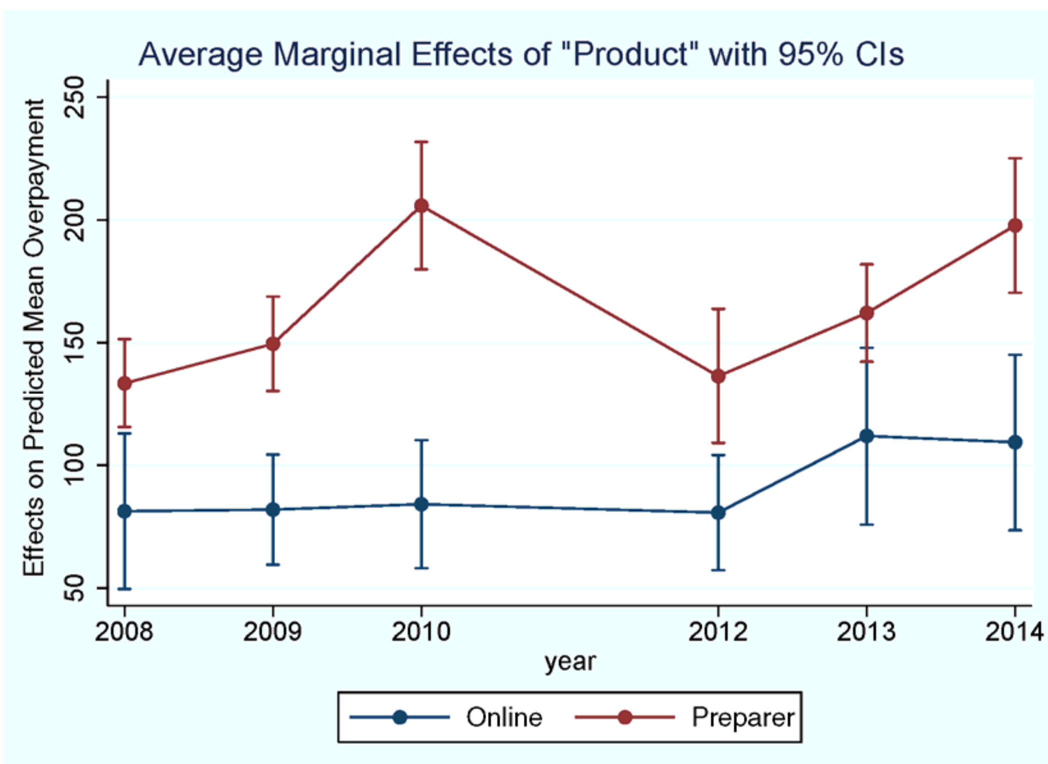

Figure 4. Average marginal effects of "product" by year and filing type. Source: Linked CPS ASEC-Form 1040 data, 2008–2010, 2012–2014, with year 2011 linearly interpolated.

Appendix

1. Further information on ineligibility calculation

Table A1 shows categories of eligibility and ineligibility by incorrect payment status and the numbers and percent of row in each category. First, 4.3 percent of those with zero earnings reported for tax purposes (either reported in the wage and salary field of the Form 1040 or reflected in a lack of W-2 earnings) erroneously filed for and received EITC. A higher-than-average rate of incorrect payment—25.6 percent—is seen for those who reported to the CPS that they had investment income above the maximum.

Table A1. Eligibility/ineligibility categories by incorrect payment status, all years

Incorrect payment	No	Yes	Total
Eligible	48,503	0	48,503
Weighted observations	110,970,354	0	110,970,354
Mean	1.00	0.00	
Zero earnings	3,241	145	5,718
Weighted observations	8,496,278	381,452	8,877,731
Mean	0.957	0.043	
Investment income over threshold	771	314	1,085
Weighted observations	1,778,040	612,941	2,390,981
Mean	0.744	0.256	
Max(income, earnings) over threshold	175,327	10,560	185,887
Weighted observations	418,680,560	25,845,908	444,526,468
Mean	0.942	0.058	
Zero dependents, age out of range	92,354	1,421	93,775
Weighted observations	216,229,599	3,552,876	219,782,475
Mean	0.984	0.016	
Other eligibility loss (via model or tax status)	1,672	1,858	3,530
Weighted observations	4,265,218	4,514,564	8,779,782
Mean	0.486	0.514	
Total	321,868	14,298	336,166
Weighted observations	760,420,049	34,907,741	795,327,791
Mean	0.956	0.044	

Source: Linked CPS ASEC-Form 1040 data, 2008–2010, 2012–2014. Cells show the number in each category of EITC ineligible tax filer and the incorrect payment rate for the type.

The cases above are straightforward—there is a clear violation of a program threshold based on a single preclusion. The largest category of ineligibles includes those whose combination of children and income lie outside the program's parameters. Filers who claimed zero dependents on the Form 1040 and were outside of the eligible age ranges had an incorrect payment rate lower than the overall average, at 1.6 percent. Filers within the age range, but with a precluding combination of income and children, received incorrect payment at a rate of 5.8 percent. Finally, a small category of filers had a conflict between the survey and tax information that appeared to make them ineligible: some filers were claimed as dependents on another filer's return; others filed as single but appeared to be married according to the household roster, thus putting their total income in question. The default stance in eligibility calculation is to accept the W-2 return information and Form 1040 values as "true" unless there is strong evidence against it (such as in the case of a spouse in the CPS ASEC household roster). This last group of filers had the highest rate of incorrect payment, at 51.4 percent, but were small in number, being less than one percent of the tax-filing population.

In contrast to work by McCubbin (2000) and Liebman (2000), which indicate that incorrect payment tends to occur through confusion or fraud regarding qualifying children, it appears as though higher rates of incorrect payment dependent on the type of potential error are associated with some of the lesser-known parameters of the program, such as the investment income threshold. The overall incorrect payment rate per taxpayer is higher for paid preparers, at 4.8 percent, than for online filers (4.2 percent) and paper filers (1.7 percent).

Table A2 shows the breakdown of these errors based on whether a filer used a tax preparer, filed online, or filed a paper form, and the incorrect payment rate for each type of error. Incorrect payment rates were about average for those reporting zero earnings, with about 4.8 percent of zerp earners filing with a tax preparer receiving incorrect payment, 3.3 percent of online filers, and 4.5 percent of paper filers. For those with misreports of investment income, online filers had highest rate of incorrect payment, at 36.0 percent. Interestingly, those who were ineligible based on their combination of income and qualifying children had rates of incorrect payment for each type of filer that were slightly above average, with those receiving preparation assistance at 6.6 percent and those filing online at 5.2 percent. The group with no dependents and appropriate earnings and income, but whose ages were out of range, had the lowest rates of all, at slightly less than 2.0 percent for both prepared and online filings. Finally, the "other" category had the highest incorrect payment rates, at 57.7 and 46.3 for prepared and online, respectively.

Table A2. Rate of incorrect payment by lack of eligibility type and filing method, all years

Filing method	Prepared	Online	Paper
Zero earnings	0.048	0.033	0.045
Total filers by method	1,948	1,004	434
Weighted filers by method	5,075,157	2,671,069	1,131,504
Investment income over threshold	0.231	0.360	0.129
Total filers by method	749	269	67
Weighted filers by method	1,627,357	601,910	161,714
Max(income, earnings) over threshold	0.066	0.052	0.017
Total filers by method	102,545	73,436	9,906
Weighted filers by method	244,352,592	176,079,508	24,094,367
Zero dependents, age out of range	0.017	0.018	0.002
Total filers by method	55,424	29,455	8,895
Weighted filers by method	130,265,031	70,083,887	19,433,557
Other eligibility loss (via model or tax status)	0.577	0.463	0.221
Total filers by method	2,055	1,246	229
Weighted filers by method	5,107,379	3,136,773	535,631
Total incorrect payment rate	0.048	0.042	0.017

Source: Linked CPS ASEC-Form 1040 data, 2008–2010, 2012–2014. Cells show the number in each category of EITC ineligible filer by filing method and the incorrect payment rate for the type.

2. Reweighting and comparison to original 1040 records

Because Form 1040 observations receive PIKs close to 100 percent of the time, calculating the probability that a CPS ASEC respondent receives a PIK amounts to calculating the probability that a tax-filing CPS ASEC respondent is found in the Form 1040 data. Table A3 compares the number of filers from each year's Form 1040 file, which can be considered the population of filers, with the weighted number of filers found in the CPS ASEC after reweighting.

Table A3: Comparison of Form 1040 counts and weighted CPS ASEC persons linked to Form 1040 data, by year			
Year	**1040 filers**	**CPS ASEC weighted filers**	**Proportion covered**
2008	138,833,790	130,203,505	0.938
2009	136,554,347	127,939,439	0.937
2010	139,057,456	127,430,267	0.916
2012	141,491,343	131,635,761	0.930
2013	142,906,142	130,800,083	0.915
2014	144,417,573	131,308,372	0.909

Source: Linked CPS ASEC-Form 1040 data, 2008–2010, 2012–2014.

Once the weights are recalculated, the number of CPS ASEC tax filers is between 91 and 94 percent of total tax filers in the 1040 file. There are several reasons why we might not cover the total number of tax filers. First, CPS ASEC respondents with imputed income data are not included in the EITC eligibility and take-up calculations. Essentially, with several variables that are key to estimation, a survey value is necessary (an example is investment income). The second issue is that Form 1040 filers may file from outside of the country. Estimates of Americans living and working abroad range from 2.2 to 6.8 million people (Costanzo and Klekowski von Koppenfels, 2013).

The statistics can be compared with an analysis of 2010 Census data (the decennial), which was linked by all persons in the Form 1040 (primary, secondary, and dependent PIKs). About 90 percent of all Form 1040 persons were found in 2010 decennial data. In the Form 1040-decennial data match, about 74 percent were White alone, about 12 percent were Black alone, and about 5 percent were Asian (which compares with 81 percent, 11 percent, and 5 percent in the present study). This higher rate for White alone may be due to primary filer characteristics differing from all persons, or due to differences in question wording between the decennial and CPS ASEC, or both. About 15 percent of persons in the Form 1040-decennial match were of Hispanic origin, compared with 14 percent in this study. Race and Hispanic origin were the only variables that were examined in the two studies.

3. Full regression results

Table A4 and A5 show the full results of the main regressions reported in the paper. The only excluded variables are the state fixed effects.

Table A4. Results of difference analysis, weighted: dependent variable is the probability of incorrect payment

	(1) Baseline	(2) With covariates	(3) Online as comparison	(4) Low income
Preparer	0.020***	0.011***	0.002	0.010***
	(0.003)	(0.002)	(0.003)	(0.003)
Online	0.018***	0.009***		0.012**
	(0.003)	(0.003)		(0.004)
Preparation X Product	0.209***	0.134***	0.083***	0.128***
	(0.013)	(0.012)	(0.011)	(0.012)
Online X Product	0.078***	0.050***		0.056***
	(0.012)	(0.011)		(0.014)
Preparer X 2009	0.004	0.004	-0.004	0.009
	(0.005)	(0.004)	(0.004)	(0.005)
Preparer X 2010	0.003	0.000	-0.019***	0.003
	(0.003)	(0.004)	(0.004)	(0.006)
Preparer X 2012	0.001	-0.003	-0.006	-0.001
	(0.004)	(0.004)	(0.003)	(0.005)
Preparer X 2013	0.011***	0.004	-0.009**	0.008*
	(0.003)	(0.003)	(0.003)	(0.004)
Preparer X 2014	0.005	-0.003	-0.002	0.003
	(0.004)	(0.005)	(0.003)	(0.007)
Online X 2009	0.009*	0.008*		0.013*
	(0.004)	(0.004)		(0.005)
Online X 2010	0.023***	0.019***		0.028***
	(0.004)	(0.005)		(0.006)
Online X 2012	0.008	0.003		0.005
	(0.004)	(0.004)		(0.006)
Online X 2013	0.021***	0.014***		0.021***
	(0.004)	(0.004)		(0.006)
Online X 2014	0.009*	-0.001		0.002
	(0.004)	(0.004)		(0.007)
Preparer X Product X 2009	0.002	0.002	-0.002	-0.002
	(0.011)	(0.011)	(0.013)	(0.012)
Preparer X Product X 2010	0.060***	0.062***	0.032	0.061***
	(0.009)	(0.009)	(0.017)	(0.010)
Preparer X Product X 2012	-0.018*	-0.021*	-0.018	-0.022*
	(0.008)	(0.008)	(0.014)	(0.009)
Preparer X Product X 2013	-0.019	-0.018	-0.033	-0.016
	(0.013)	(0.012)	(0.017)	(0.013)
Preparer X Product X 2014	-0.002	0.001	-0.015	-0.000
	(0.012)	(0.011)	(0.016)	(0.011)
Online X Product X 2009	0.002	0.004		0.002
	(0.012)	(0.011)		(0.015)
Online X Product X 2010	0.029*	0.030*		0.030
	(0.013)	(0.013)		(0.016)
Online X Product X 2012	0.002	-0.003		-0.008
	(0.012)	(0.012)		(0.016)
Online X Product X 2013	0.018	0.014		0.013

	(0.013)	(0.013)		(0.016)
Online X Product X 2014	0.015	0.016		0.020
	(0.012)	(0.011)		(0.014)
Product			0.050***	
			(0.011)	
Product X 2009			0.004	
			(0.011)	
Product X 2010			0.030*	
			(0.013)	
Product X 2012			-0.003	
			(0.012)	
Product X 2013			0.014	
			(0.013)	
Product X 2014			0.016	
			(0.011)	
Year=2009	0.003	0.001	0.009**	0.000
	(0.003)	(0.003)	(0.003)	(0.003)
Year=2010	0.008*	0.010*	0.030***	0.013*
	(0.003)	(0.004)	(0.004)	(0.006)
Year=2012	0.002	0.005	0.008**	0.004
	(0.003)	(0.003)	(0.003)	(0.004)
Year=2013	-0.004	0.003	0.016***	0.000
	(0.004)	(0.004)	(0.003)	(0.005)
Year=2014	0.002	0.012*	0.012***	0.011
	(0.005)	(0.005)	(0.002)	(0.007)
Age		0.006***	0.006***	0.009***
		(0.000)	(0.000)	(0.000)
Age squared		-0.000***	-0.000***	-0.000***
		(0.000)	(0.000)	(0.000)
Sex		-0.039***	-0.042***	-0.036***
		(0.004)	(0.005)	(0.004)
Head of household		-0.022***	-0.025***	-0.070***
		(0.006)	(0.006)	(0.006)
Married		0.023***	0.024***	0.025***
		(0.003)	(0.003)	(0.003)
One child		0.073***	0.073***	0.134***
		(0.004)	(0.004)	(0.004)
Two children		0.063***	0.063***	0.125***
		(0.003)	(0.002)	(0.003)
Three or more children		0.042***	0.042***	0.089***
		(0.003)	(0.003)	(0.007)
Black		0.068***	0.070***	0.075***
		(0.007)	(0.007)	(0.007)
Asian		0.007*	0.008**	0.017***
		(0.003)	(0.003)	(0.004)
Other		0.009*	0.009*	0.011*
		(0.004)	(0.004)	(0.005)
Hispanic		0.043***	0.043***	0.040***

		(0.003)	(0.003)	(0.003)
Native-born citizen		0.021***	0.021***	0.016***
		(0.002)	(0.002)	(0.003)
HS degree		-0.006*	-0.007*	-0.005
		(0.003)	(0.003)	(0.003)
Some college		-0.024***	-0.025***	-0.022***
		(0.003)	(0.003)	(0.003)
BA/BS+		-0.035***	-0.037***	-0.035***
		(0.004)	(0.004)	(0.004)
Self-employed		0.131***	0.122***	0.091***
		(0.015)	(0.016)	(0.017)
Income zero or less		0.067***	0.067***	0.056***
		(0.003)	(0.003)	(0.004)
Log income		-0.005***	-0.005***	-0.004***
		(0.000)	(0.000)	(0.000)
Log income squared		0.107***	0.109***	0.153***
		(0.005)	(0.005)	(0.007)
No TANF		0.030	0.022	0.035
		(0.092)	(0.093)	(0.093)
Log TANF value		0.005	0.004	0.006
		(0.012)	(0.012)	(0.012)
No SNAP		-0.021	-0.025	-0.022
		(0.013)	(0.014)	(0.013)
Log SNAP value		0.004	0.004	0.000
		(0.002)	(0.003)	(0.003)
Urban		-0.013***	-0.012***	-0.011**
		(0.002)	(0.002)	(0.003)
Constant	-0.004	-0.310**	-0.286**	-0.326***
	(0.003)	(0.089)	(0.090)	(0.087)
N	336,166	336,166	313,632	211,490

* p<0.05, ** p<0.01, *** p<0.001. Source: Linked CPS ASEC-Form 1040 data, 2008-2010, 2012-2014. Results from a difference-in-differences model comparing five groups: paper filers as the base category; preparer with a product; online with a product; and preparer and online with no product.. Standard errors clustered at the state level shown in parentheses.

Table A5. Results of difference analysis, GLM with a log link, weighted: dependent variable is the value of overpayment	(1) Baseline	(2) With covariates	(3) Online as comparison
Preparer	0.836***	0.663***	0.027
	(0.122)	(0.126)	(0.072)
Online	0.823***	0.641***	
	(0.131)	(0.131)	
Preparation X Product	1.861***	0.885***	0.171
	(0.082)	(0.055)	(0.113)
Online X Product	1.173***	0.714***	
	(0.128)	(0.122)	
Preparer X 2009	0.070	0.056	-0.041
	(0.162)	(0.150)	(0.110)
Preparer X 2010	0.100	0.012	-0.247*
	(0.166)	(0.185)	(0.097)
Preparer X 2012	-0.179	-0.304	-0.256*
	(0.225)	(0.218)	(0.102)
Preparer X 2013	0.246	0.005	-0.192**
	(0.191)	(0.201)	(0.065)
Preparer X 2014	-0.036	-0.280	-0.127
	(0.284)	(0.286)	(0.104)
Online X 2009	0.097	0.097	
	(0.162)	(0.150)	
Online X 2010	0.351	0.258	
	(0.184)	(0.204)	
Online X 2012	0.066	-0.049	
	(0.228)	(0.225)	
Online X 2013	0.416*	0.197	
	(0.198)	(0.203)	
Online X 2014	0.090	-0.153	
	(0.262)	(0.267)	
Preparer X Product X 2009	-0.041	0.007	0.075
	(0.073)	(0.063)	(0.144)
Preparer X Product X 2010	0.050	0.117	0.323*
	(0.079)	(0.072)	(0.157)
Preparer X Product X 2012	-0.094	-0.053	0.121
	(0.089)	(0.095)	(0.195)
Preparer X Product X 2013	-0.146	-0.032	0.028
	(0.079)	(0.066)	(0.155)
Preparer X Product X 2014	-0.075	0.047	0.120
	(0.079)	(0.073)	(0.157)
Online X Product X 2009	-0.095	-0.068	
	(0.141)	(0.140)	
Online X Product X 2010	-0.228	-0.205	
	(0.154)	(0.138)	
Online X Product X 2012	-0.129	-0.174	
	(0.173)	(0.172)	
Online X Product X 2013	-0.067	-0.062	

	(0.146)	(0.147)	
Online X Product X 2014	-0.094	-0.073	
	(0.157)	(0.145)	
Product			0.719***
			(0.122)
Product X 2009			-0.070
			(0.140)
Product X 2010			-0.207
			(0.138)
Product X 2012			-0.175
			(0.172)
Product X 2013			-0.063
			(0.147)
Product X 2014			-0.075
			(0.146)
Year=2009	0.082	0.048	0.145*
	(0.123)	(0.114)	(0.073)
Year=2010	0.159	0.232	0.492***
	(0.174)	(0.190)	(0.074)
Year=2012	0.317	0.414*	0.365***
	(0.195)	(0.194)	(0.085)
Year=2013	0.067	0.251	0.448***
	(0.203)	(0.215)	(0.059)
Year=2014	0.361	0.597*	0.444***
	(0.276)	(0.283)	(0.086)
Age		0.167***	0.165***
		(0.007)	(0.007)
Age squared		-0.002***	-0.002***
		(0.000)	(0.000)
Sex		-0.812***	-0.816***
		(0.032)	(0.033)
Head of household		-0.928***	-0.935***
		(0.040)	(0.039)
Married		0.233***	0.229***
		(0.027)	(0.027)
One child		0.710***	0.694***
		(0.035)	(0.036)
Two children		0.780***	0.765***
		(0.041)	(0.043)
Three or more children		0.684***	0.675***
		(0.057)	(0.058)
Black		0.860***	0.854***
		(0.048)	(0.049)
Asian		0.365***	0.374***
		(0.038)	(0.039)
Other		0.226***	0.227***
		(0.057)	(0.057)
Hispanic		0.606***	0.599***

		(0.052)	(0.052)
Native-born citizen		0.177***	0.177***
		(0.052)	(0.052)
HS degree		-0.038	-0.038
		(0.021)	(0.021)
Some college		-0.341***	-0.343***
		(0.018)	(0.019)
BA/BS+		-0.862***	-0.859***
		(0.042)	(0.043)
Self-employed		12.485***	12.397***
		(1.617)	(1.628)
Income zero or less		3.234***	3.208***
		(0.343)	(0.348)
Log income		-0.184***	-0.183***
		(0.018)	(0.019)
Log income squared		0.950***	0.938***
		(0.036)	(0.038)
No TANF		0.057	0.022
		(0.627)	(0.619)
Log TANF value		0.024	0.020
		(0.080)	(0.079)
No SNAP		-0.153	-0.147
		(0.112)	(0.113)
Log SNAP value		-0.018	-0.018
		(0.021)	(0.022)
Urban		-0.142***	-0.115***
		(0.035)	(0.034)
Constant	2.702***	-13.000***	-12.179***
	(0.116)	(2.013)	(1.989)
N	336,166	336,166	313,632

* p<0.05, ** p<0.01, *** p<0.001. Source: Linked CPS ASEC-Form 1040 data, 2008-2010, 2012-2014. Results from a difference-in-differences model comparing five groups: paper filers as the base category; preparer with a product; online with a product; and preparer and online with no product. Standard errors clustered at the state level shown in parentheses.