ELSEVIER

Journal of Public Economics 91 (2007) 1351–1367

JOURNAL OF PUBLIC ECONOMICS

www.elsevier.com/locate/econbase

Electronic filing, tax preparers and participation in the Earned Income Tax Credit ☆

Wojciech Kopczuk [a,*], Cristian Pop-Eleches [b]

[a] *Columbia University and NBER, United States*
[b] *Columbia University, United States*

Received 22 November 2005; received in revised form 10 November 2006; accepted 14 November 2006
Available online 2 January 2007

Abstract

In 2002 more than 18 million low-income individual taxpayers received the Earned Income Tax Credit (EITC). Despite its size, non-participation in this program is a concern and substantial effort is devoted by the IRS, local governments and many non-profits to address it. Using variation across states in the introduction of state electronic filing programs, we show that the introduction of electronic filing had a significant effect on participation in the EITC. Our results are robust to accounting for other welfare, EITC and IRS reforms introduced during the same period. We speculate that the impact of this policy change on the tax preparation industry played an important role in increasing participation.

Keywords: Take up of social benefits; Tax preparers; Earned Income Tax Credit

1. Introduction

In a recent review of the literature on the take-up of social programs, Currie (2004) concludes that despite years of research effort, there is relatively little knowledge about the main barriers to increased program participation as well as the best policies to decrease these barriers. In this paper we focus on an important social program, the Earned Income Tax Credit (EITC), and study how the introduction of a particular technology (the ability to file taxes electronically) affected the number of EITC claims.

☆ We are grateful to anonymous referees, Michael Conlin, Gary Engelhardt, Brendan O'Flaherty, Jeff Liebman, Emmanuel Saez, Lucie Schmidt, Till von Wachter and seminar participants at Columbia, Syracuse, Harvard/MIT and Duke for helpful comments.

* Corresponding author.
E-mail addresses: wkopczuk@nber.org (W. Kopczuk), cp2124@columbia.edu (C. Pop-Eleches).

0047-2727/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpubeco.2006.11.003

More specifically, we use the staggered introduction of state electronic filing programs during the 1990s to show that these state programs had a large effect on the number of federal electronic filers as well as the number of EITC claims. Our main identifying assumption is that once we control for state and year effects as well as state-specific time trends, the timing of the introduction of the state electronic filing programs is not correlated with other factors that might have an effect on our outcome variables.

We carry out our analysis using annual public use samples of federal income tax returns issued by the IRS for the period 1988 to 1999 to generate yearly measures of electronic filing and EITC participation by state and income categories. These data are complemented with information from the IRS on the date of the introduction of electronic filing programs for state income taxes.

We find evidence that in an average state, the introduction of state electronic filing increases the number of federal returns that are filed electronically by an additional .6 to 1.1 percentage points. Next we show a similarly large increase in the number of people claiming EITC on their federal tax returns. Our instrumental variable estimate implies that 79% of the increase in electronic filing that resulted from the state electronic filing programs was due to new EITC participants. To test the robustness of our main results, we present a number of specification checks that try to account for possible competing alternative explanations of our results, such as the welfare reform and state-specific welfare waivers implemented during this period, or IRS programs to reduce EITC related cheating.

Our analysis proceeds as follows. We describe the EITC program and the empirical strategy in Section 2. In Section 3, we introduce our data and regression framework. The main empirical analysis is presented in Section 4. We conclude in Section 5 with the discussion of possible mechanisms that might explain our results. In particular, we highlight the role of the tax preparation industry as well as the incentives implicitly provided by the IRS for the industry to target the EITC population.

2. Background and empirical strategy

2.1. EITC

The history and research findings on the EITC have recently been reviewed in a survey article by Hotz and Scholz (2003). Therefore, we will only focus on the features of the program that are relevant for our analysis. The EITC, which is administered through the tax system, was expanded between 1991 and 1994 both by increasing the maximum benefits by almost 50% in real terms and by extending eligibility to childless individuals. The 1999 (the last year we use in our analysis) maximum for workers with two or more children was $3816, with the phase in range between $0 and $9540 and the phase out range beginning at $12,460. The credit was fully phased out at $30,580. For single individuals the maximum benefit was just $347 in 1999 and the program was fully phased out when income reached $10,200.[1]

Although the basic computation of the EITC is straightforward, the eligibility criteria are not. The major difficulty lies in determining whether the taxpayer meets the qualifying child criteria, since age, disability and residency restrictions are in place. Apart from the criteria related to children, there is also room for manipulating the system by adjusting the filing status. For

[1] In 2002, eighteen states had their own EITC programs. This is a relatively new phenomena: in 1999 there were twelve such states, and in 1990 only five. We ignore state EITC programs in most of the paper and provide only some robustness checks that indicate that these programs do not affect our analysis.

example, for some income levels a married couple that files separately as two heads of household could receive a larger credit than when filing jointly. Finally, another source of potential fraud and complexity within the system has to do with the definition of income for EITC, since certain types of income (such as interest, dividend, rent and royalties and capital gains) can make taxpayer ineligible if they exceed the threshold amount ($2350 in 1999). There is evidence of non-compliance related to the eligibility criteria (McCubbin, 2000; Liebman, 2000), but also some evidence that many mistakes are not intentional. Using a random sample of EITC recipients examined by the IRS's Criminal Investigations Division, McCubbin (2000) finds that most of the overclaims are due to qualifying child errors (almost 70% of the total amount overclaimed), but the auditors believed that approximately 50% of errors were not intentional (McCubbin, 2000).

2.2. Non-participation in EITC

Apart from the issue of non-compliance, non-participation in the EITC program has received significant attention from the government and researchers. In an older study, Scholz (1994) estimated that in 1990 approximately 80% of eligibles actually claimed benefits. A more recent IRS study (Internal Revenue Service, 2002) states that in 1996 between 13% (based on march the CPS) and 18% (based on SIPP) eligible individuals (between 2.3 and 3.4 million) did not file tax returns and therefore did not take advantage of the program.[2]

There are a number of factors that might have contributed to an increase in EITC participation rates, conditional on eligibility, during our period of study. First, the EITC expanded to include richer families, who could have larger participation rates. At the same time both governmental and non-governmental organizations have devoted resources to outreach activities to increase participation. In addition, the increase in the value of the credit might have induced increased participation among some eligibles.

On the other hand, Hotz and Scholz (2003) speculate that non-participation in the 1990s might have increased for some groups due to changes in labor market participation of women as well as repeated expansions during this period that forced taxpayers to learn new rules. Hill et al. (1999) provide evidence suggesting that participation rates among women who used to be on welfare are much lower (even as low as 50%) than overall participation rates. The continued increase in the share of Hispanics in the US population might also have affected the rate of non-participation in recent years. Recent survey evidence (Maag, 2005) indicates that among Hispanic low-income parents, only 27% have heard of the EITC, compared to an average of 58% for the entire population. Finally, non-participation in the EITC is also affected by the existence of a significant pool of non-filers of federal taxes. In a recent paper, Erard and Ho (2001) calculate that in 1988 about $5 billion in unpaid taxes are due to people who are legally obliged to file but do not do so (the so-called "ghosts"). Most interesting in our context is their estimate that 29% of "ghosts" might actually be entitled to get money back but somehow do not file and, moreover, that "ghosts" are likely to be over-represented in the EITC eligible groups.

[2] There are a number of reasons why these numbers might be an overestimate of the true non-participation rates during our period of study. First, IRS's computations rely on a very questionable treatment of a large group of potentially eligible individuals. They include in the denominator (therefore treating them as if they were eligible) 23% of otherwise eligible individuals who refused to provide their SSN in the CPS and 7% of respondents in the SIPP who refused to answer the question about filing a tax return. However these same individuals are not counted as non-participants. Both estimates also assume that an eligible taxpayer who filed a tax return actually claimed benefits. Holtzblatt and McCubbin (2003) present back-of-the-envelope calculations suggesting that the true non-participation rate probably exceeds 20%.

2.3. Empirical strategy

The effect of electronic filing on EITC participation is identified by the introduction of state electronic filing programs during the 1990s.[3] The IRS introduced electronic filing of federal tax returns in 1986, but states started introducing their own electronic filing systems for the payment of state income taxes in the early 1990s. Most income tax states (except for California, Maine, Massachusetts and Minnesota) took advantage of the broader IRS initiative operated by the IRS Office of FedState Relations established in 1991.[4] In these states, there is only one electronic return to be filed, with the IRS transmitting the relevant information to the state on the taxpayers' behalf. We use this differential timing to show that the introduction of state electronic filing had a significant effect on the number of federal returns that were filed electronically and then we look at the effect on EITC participation.

We use information provided by the IRS to code the existence of a program allowing electronic filing of state income taxes and to date the introduction of such programs.[5] Table 1a shows the timing of introduction of the electronic filing programs. In almost all cases, in the year of the introduction the program was tested on a small population and, therefore, our treatment variable is equal to one if the given year is *strictly greater* than the year shown in the table.[6] There is substantial variation in the timing of state e-filing programs. Minnesota introduced its program in 1989, while South Carolina, Maryland and Illinois did so in 1990 (of these, only South Carolina's program was a part of the Federal/State Electronic Filing Program). Another 4 states implemented the program in 1991, 10 in 1992, 8 in 1993 and 7 in 1994. Implementation in the remaining nine states was spread over the following six years, with Vermont and Hawaii adopting in 2001.[7]

Our key identifying assumption is that once we include controls, the timing of the introduction of the state electronic filing programs is not correlated with other factors that might have an effect on our outcome variables. Fig. 1 shows estimates of non-participation rates in 1996 as estimated by the Internal Revenue Service (2002) against the timing of the introduction of the state electronic filing programs. There is no obvious relationship between the two variables and there is also very significant variation in non-participation rates suggesting that the non-participation rates were unrelated to the decision to introduce a state e-filing program. The non-participation rate for states treated pre-1996 should be affected due to early introduction of the programs, obscuring the interpretation of this picture. However the remaining large variation in the non-participation rates suggests that the pre-existing differences are not likely to explain the timing of the programs.

[3] Most electronic filing is done by professionals who are authorized by the IRS as Electronic Return Originators (ERO). Some taxpayers choose to prepare their own tax return and have a professional file it electronically. In other cases, a paid tax professional also prepares the tax return. In the late 1990s, many taxpayers chose to file their tax returns by phone under the TeleFiling program, but it was not possible to claim EITC via Telefile. Finally, the direct on-line filing is another possibility to file electronically, but its importance was small in the period that we study. In 1999, the last year in our sample, approximately 8% of electronic returns were filed online and the penetration of this method was most likely even lower at the bottom of the income distribution. Our data does not allow for separating online filing from other means of electronic filing.

[4] This was one of the IRS's efforts intended to achieve the 80% electronic filing rate by 2007, as outlined in the IRS Restructuring and Reform Act of 1998.

[5] http://www.irs.gov/pub/irs-utl/state2.pdf, accessed December 15, 2003.

[6] The table also shows the year of introduction of state TeleFiling programs and the timing of joint Federal/State Programs.

[7] As a robustness check, we also perform our analysis by defining all states without a state income tax as never treated. Alternatively, we also define our treatment as the introduction of the *Federal/State Electronic Filing Program* since this program is arguably uniformly applied across states.

Table 1a
Summary statistics, by state (1988–1999)

	Year of			% of returns in	% of returns with		
	State EFI	Telefile	StateFed	0–30K	EITC	E-file	Professional help
AL	1997	None	1997	63.78	23.56	18.71	52.58
AK	No tax	No tax	No tax	63.85	8.06	8.93	40.03
AZ	1997	None	1997	60.00	16.45	12.69	53.32
AR	1994	1997	1994	66.26	21.57	16.72	57.33
CA	1994	1996	Never	55.35	15.22	7.88	57.00
CO	1993	None	1993	53.27	11.15	11.05	50.09
CT	1993	1998	1993	50.85	7.69	9.95	46.04
DE	1993	1999	1993	51.54	11.39	12.04	42.40
DC	1996	2000	1996	57.20	14.13	12.20	42.21
FL	No tax	No tax	No tax	61.59	15.52	15.80	50.03
GA	1994	2001	1994	58.42	19.05	18.58	51.41
HI	2001	None	2001	56.59	8.07	7.18	53.31
ID	1993	1999	1993	66.81	17.31	12.11	48.05
IL	1990	1993	1995	55.57	12.18	14.44	51.25
IN	1992	1999	1992	56.26	11.76	18.33	50.44
IA	1993	None	1993	59.18	9.00	13.30	62.74
KS	1991	1996	1991	56.94	12.18	14.15	57.94
KY	1992	1999	1992	61.40	15.53	17.75	55.68
LA	1992	1996	1992	63.42	24.75	17.90	49.72
ME	2000	1996	Never	61.32	10.03	10.82	39.04
MD	1990	2001	1995	52.24	11.23	11.19	44.98
MA	1992	1995	Never	53.05	7.41	9.58	47.48
MI	1992	1999	1992	55.44	10.87	12.51	50.03
MN	1989	1998	Never	56.26	7.87	10.70	50.85
MS	1992	1995	1992	68.30	30.53	20.27	47.46
MO	1993	1997	1993	60.50	14.59	15.17	56.55
MT	1994	1997	1994	65.88	13.59	13.08	57.38
NE	1993	1998	1993	58.47	12.05	10.67	57.30
NV	No tax	No tax	No tax	57.71	13.80	13.56	50.24
NH	No tax	No tax	No tax	52.96	7.71	13.00	40.62
NJ	1994	1996	1994	49.69	10.20	9.48	53.77
NM	1992	None	1992	64.99	24.44	16.88	49.66
NY	1992	None	1992	54.97	12.59	8.76	55.78
NC	1991	None	1991	61.48	17.08	17.48	51.89
ND	1999	None	1999	59.92	9.15	10.71	60.88
OH	1999	1997	1999	59.13	10.72	14.01	48.11
OK	1992	2001	1992	63.09	16.97	14.30	57.54
OR	1993	2001	1993	55.42	10.16	8.22	47.71
PA	1995	1998	1995	58.65	10.56	10.28	45.35
RI	1994	None	1994	54.94	11.31	10.35	50.88
SC	1990	1997	1990	65.62	19.78	23.78	53.33
SD	No tax	No tax	No tax	66.00	11.92	14.99	50.58
TN	No tax	No tax	No tax	62.13	18.57	20.01	49.23
TX	No tax	No tax	No tax	61.63	19.94	16.67	45.33
UT	1992	1998	1992	56.46	11.22	11.90	45.62
VT	2001	None	2001	57.32	10.79	6.72	43.37
VA	1994	2000	1994	53.25	11.27	13.29	43.85
WA	No tax	No tax	No tax	51.88	9.24	11.50	42.33
WV	1991	2001	1991	64.49	17.04	14.57	42.33
WI	1991	1998	1991	57.93	9.71	12.63	48.97
WY	No tax	No tax	No tax	60.83	9.59	16.07	50.89
Total				58.83	13.94	13.07	50.84

Notes: The first three columns refer to the date of the introduction of the first state electronic filing program (State EFI), the state telefiling program (Telefile) and the IRS sponsored Federal/State Electronic Filing Program (StateFed).

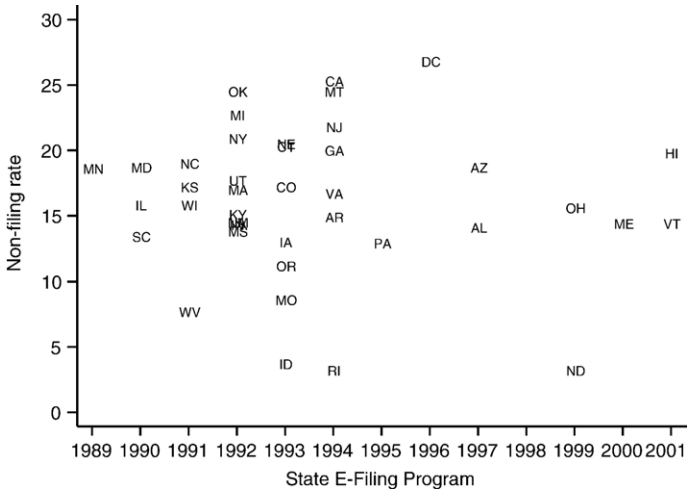

Fig. 1. Non-participation rate. *Notes*: The date of the introduction of a state e-filing program is taken from Table 1a. The non-participation rate is for 1996 and taken from the Internal Revenue Service (2002).

Similarly, we plot in Fig. 2 the average number of EITC recipients by state in 1989 against the timing of the state electronic filing programs and also find no visible relation between these two variables.

3. Data and regression framework

Our main sources of data are the annual public use samples of federal income tax returns ("Tax Model Files") issued by the Statistics of Income (SOI) division of the IRS and available from the NBER.[8] Our study period starts in 1988, the first year for which our dataset includes information on whether a tax return was filed electronically, and stops in 1999, the latest year for which the data are available. The size of yearly samples varies between 90,0000 and 130,000 observations, with between 13,000 and 32,000 observations having adjusted gross income below 30,000.[9] The public use files contain most of the information from federal tax returns, such as the state of residence, the EITC status and payment, information on whether a professional was involved in preparing the return, the electronic filing indicator, adjusted gross income, filing status and the number of dependents.

We aggregate all information by state and income categories. The main reason for this approach is that we doubt the validity of the assumption that our samples are identical from one year to the next. In particular, the natural channel of any response in participation is due to an increase in the number of filers. Secondly, the existence of an electronic filing program varies only by year and state so that our source of identification is not compromised by aggregation. Over the twelve year long period (1988–1999) the fifty states plus the District of Columbia result in 612 cells. We usually use fewer observations because in our main specifications we are not using states without their own income tax.

[8] The documentation for this dataset can be accessed at www.nber.org/taxsim.
[9] In this dataset high income individuals are oversampled.

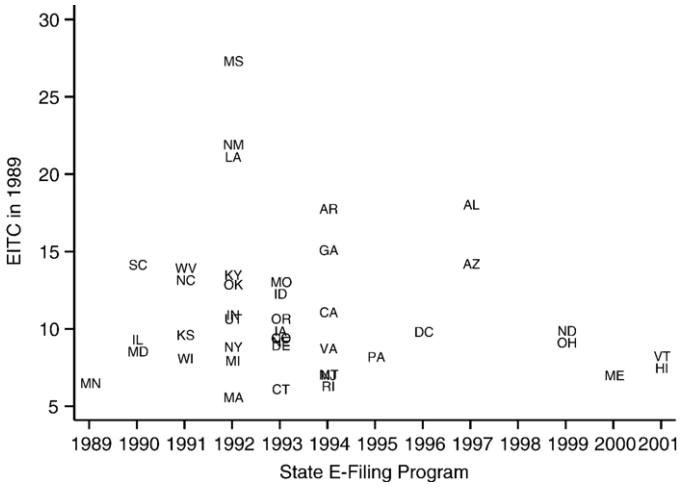

Fig. 2. EITC in 1989. *Notes*: The date of the introduction of a state e-filing program is taken from Table 1a. The percent of EITC recipients in 1989 is based on the author's calculations based on the Tax Model Files issued by the Statistics of Income division of the IRS.

We use the adjusted gross income to divide taxpayers into categories. Both adjusted gross income and earned income have to be below the relevant threshold to be eligible for the EITC.[10] Since the maximum threshold between 1988 and 1999 was $31,794 (in real 2000 dollars), all our income categories are defined to be below this maximum threshold.[11]

In addition, in our main specifications we also restrict ourselves to households above $10,597, and refer to this category as "10–30K" group. The reason for using this threshold, rather than all the households below the maximum threshold of $31,794, is to better account for changes in the criteria for receiving the EITC as a result of the 1994 expansion of the credit to childless individuals with incomes below $10,597.[12] By defining this "10–30K" income category that excludes childless EITC recipients, we can isolate the impact of these 1994 changes in eligibility criteria on the number of EITC recipients. We also exclude roughly 5% of individuals in this category who received EITC but did not claim any child exemptions.[13] This effect is possibly large and in our data we cannot precisely distinguish between different categories of EITC filers since we do not have information about the number of children that are qualifying for EITC purposes. However, we also present some results for the whole category of individuals below $31,794 ("0–30K").[14] We also decompose our results further based on the $0–$10,597 category

[10] The dataset contains a measure of earned income, constructed by the SOI. Although taxpayers do not have to report their Earned Income directly, they need to determine it for the purpose of establishing eligibility and computation of the credit on the so-called EIC form. This form does not have to be submitted with the tax returns. We found that the constructed earned income measure does not match closely with the size of the EITC refund. It is also not available for non-EITC recipients. The AGI does not suffer from these problems.

[11] All our dollar figures are expressed in 2000 dollars and adjusted for inflation.

[12] More precisely, the maximum income for receiving these benefits as a childless claimant varied between $10,464 and $10,596 during this period.

[13] The results in the working paper version (Kopczuk and Pop-Eleches, 2005) included those individuals in the 10–30K category, thereby explaining minor numerical differences in results. The criteria for claiming a dependent exemption and for claiming a child for EITC purposes are different. The presence of these individuals is evidence that exemptions and qualifying children for EITC purposes are also not the same thing in practice.

[14] All EITC recipients who claimed at least one dependent exemption are included in this last category.

Table 1b
EITC, e-filers and professionally prepared returns by income group and year

	% EITC		% e-filers			% professionally prepared return			% e-filers and professional returns
	0–30K	10–30K	0–30K	>30K	EITC	0–30K	>30K	EITC	EITC
1988	0.18	0.21	0.01	0.01	0.04	0.41	0.57	0.43	0.03
1989	0.18	0.21	0.04	0.03	0.13				
1990	0.19	0.22	0.09	0.06	0.25				
1991	0.21	0.23	0.12	0.06	0.36	0.45	0.55	0.51	0.23
1992	0.22	0.24	0.14	0.07	0.40	0.46	0.55	0.55	0.27
1993	0.23	0.25	0.16	0.08	0.42	0.46	0.55	0.55	0.29
1994	0.28	0.25	0.13	0.07	0.26	0.46	0.55	0.52	0.19
1995	0.28	0.26	0.16	0.08	0.29	0.46	0.56	0.54	0.21
1996	0.28	0.26	0.20	0.10	0.36	0.47	0.56	0.57	0.28
1997	0.28	0.26	0.25	0.13	0.42	0.49	0.57	0.60	0.34
1998	0.28	0.26	0.28	0.17	0.45	0.50	0.57	0.62	0.37
1999	0.27	0.26	0.32	0.23	0.54	0.52	0.58	0.67	0.45
Average	0.24	0.24	0.16	0.10	0.34	0.47	0.10	0.56	0.28

Notes: Tables are based on authors' calculations based on the Tax Model Files issued by the Statistics of Income division of the IRS.

(and call it the "0–10K" category) and divide the "10–30K" category into two brackets by splitting it at $20,000, thus creating two additional income categories (called "10–20K" and "20–30K").

3.1. Summary statistics

The first two columns of Table 1b present the mean number of people who claim EITC on their federal tax return by year and income category for the period 1988–1999. The data shows an increase in the penetration of the program during this period from 18% to 27% in the 0–30K income category (an increase from 11 to 19 million). The table also highlights the major federal expansions of the EITC program in 1993, but this effect was in fact limited to the 0–10K category. The increase in participation following the 1993 expansion is mostly due to extending eligibility to childless individuals. The total number of EITC recipients increased from 15.1 to 19 million between 1993 and 1994. In our 10–30K category (the range that excludes childless individuals), the number of EITC returns increased from 10 to 10.3 million between 1993 and 1994, while the number of returns in the 0–10K category, that includes all of the childless EITC recipients, increased from 5.1 to 8.6 million. The same table shows that although introduced in 1986, federal e-filing grew slowly in the first years after implementation so that by 1990 less than 10% of tax returns were filed electronically. However, federal electronic filing grew steadily in the 1990s.

Trends in electronic filing by income category over this period are particularly interesting. For the 0–30K category, the proportion of electronic filers is higher throughout this period indicative of the fact that poorer people are more likely to file electronically. Furthermore, EITC recipients were much more likely to file electronically than a typical low-income person. The proportion of electronic filers in the higher income groups is generally below 10% throughout this period with the exception of the period 1997–1999 when on-line filing becomes a popular method of filing.

Table 1b also summarizes the proportion of tax returns that were completed with the help of a paid tax preparer by income brackets and over time. Somewhat surprisingly, the average number of people relying on outside help in preparing their tax returns is around 50% and comparable across income groups. Perhaps this reflects the balancing of two separate effects: people in lower income groups have a harder time filling out their returns without help from preparers but people with higher incomes have more complicated returns. The lower income groups experienced the largest uptake in relying on tax professionals and the increase was particularly large among EITC recipients.

The extent of reliance on both electronic filing and professional tax services among EITC claimants is described in the last column of Table 1b. Among EITC recipients, the proportion of EITC claims submitted electronically is large, growing over time and averages about 30%. In conclusion, these summary statistics provide strong evidence that people in income groups that claim EITC on their federal tax returns are very likely to submit their tax returns electronically and usually with the help of a tax preparer. Therefore any policy intervention that would reduce the cost of filing electronically, as is the case with the introduction of electronic filing of state taxes discussed above, could potentially have an effect on the number of EITC claims.

3.2. Regression framework

We estimate OLS regressions of the following form:

$$\text{outcome}_{st} = \beta_0 + \beta_1 \text{treat}_{st} + \beta_2 \delta_s + \beta_3 \tau_t + \beta_4 t\delta_s + \epsilon_{st}, \quad (1)$$

where outcome_{st} is one of our dependent variables of interest (electronic filers or EITC claimants), measured in shares for each year and state cell in a given income category. The variable treat_{st} is a dummy variable equal to one if a state had adopted state electronic filing and δ_s and τ_t are a set of dummy variables for state and year respectively. Our preferred specification also controls for state-specific trends ($t\delta_s$), which should account for other unobservable factors that might be slowly changing at the state level over time.

In our analysis, we run a number of alternative specifications in order to test the validity of our results. We present figures based on the models that include both leads and lags in order to better understand whether we appropriately control for any trends in the data before the introduction of joint electronic filing. Since in our panel not all states have data available for each year relative to the implementation date of electronic filing of state taxes, the number of states identifying a particular coefficient is not constant and these compositional changes could give rise to possible trends in the data around the implementation date. Therefore, we also include specifications using a "balanced" panel of states that have at least 3 or 4 years of pre- and post-treatment data. In our basic analysis we concentrate on the 10–30K income category. The primary motivation for this restriction, as discussed earlier and presented in Table 1b, is that in 1994, the EITC eligibility criteria for very low-income individuals were extended and the number of EITC claims in the 0–10K category jumped discontinuously. No similar complications are present for the 10–30K category.

We define our outcome variables in shares, which are defined by dividing the number of individuals with a particular characteristic in the category of interest by *the total number of*

returns filed in the state during the given year. In other words, these are shares of all tax filers. As a result, these numbers are directly comparable and additive across income categories.[15]

Bertrand et al. (2004) have recently pointed out that due to error terms in empirical frameworks similar to ours having unknown autocorrelation structure, regular standard errors are potentially very significantly biased (usually downwards). We address this problem by reporting results from two different approaches. Our baseline standard errors are based on the standard block bootstrap procedure (using 10,000 iterations). Additionally, we report p-values from the block bootstrap procedure,[16] and the *p*-values based on the estimate of the robust covariance matrix allowing for a general form autocorrelation (i.e., clustered at the state level). Our major conclusions do not depend on the choice of the approach; due to the relatively small sample size we concentrate on bootstrapped standard errors that are usually (but not always) smaller than the (asymptotically consistent but with unknown small sample properties) clustered errors and larger than the robust ones.

4. Results: Effects on electronic filing and EITC claims

4.1. Graphical evidence

A simple way to depict the main results of the paper is to demonstrate the effect graphically. In Fig. 3a, we define for each state as year 0 the year in which state electronic filing was introduced and we plot the average fraction of federal electronic filers and EITC recipients with gross income in the 0–30K and 10–30K categories relative to the base year for a four year window. We observe increases in EITC claims following the introduction of state electronic filing. In the case of electronic filing, the series based on the raw data are heavily trending. Thus, we include the analogous series obtained from e-filing rates detrended based on the results of a year fixed effects regression. The detrended series reveal the presence of a response of electronic filing as well. Fig. 3b presents the patterns in e-filing and EITC that account for state, time and state specific time trend effects and which correspond to the regressions discussed in detail below. The pattern of the fraction of e-filers and EITC recipients is striking as it shows a discontinuous jump around year zero.

4.2. Electronic filing

In the top panel of Table 2, we show the estimated effect of the introduction of electronic filing programs for state taxes on the number of federal tax returns filed electronically estimated using

[15] As a robustness check, we also define our outcome variables in logs. This is complicated by the fact that in the beginning of the period (1988–1990) we have a small number of zeros in the sample. Therefore, in order to take advantage of the convenient percentage interpretation of logged variables, for any variable involving electronic filing we take logs of the number of individuals *plus* 1% of state tax returns. This adjustment is mechanically customized to the state's population and addresses the problem of zeros. Some changes in our variables of interest may simultaneously correspond to changes in the number of tax returns in a given state but will increase the sensitivity of our variable to such changes by at most 1%. This adjustment should not therefore affect our results in a quantitatively important way but it allows us to keep our sample balanced. The working paper version (Kopczuk and Pop-Eleches, 2005) contains results using both log and share measures of EITC participation. We also estimated all our specifications using state population as the denominator and found no qualitative differences in the results.

[16] *P*-values are based on bootstrapping the studentized coefficient on the treatment dummy. As discussed by Horowitz (2000), if possible, one should bootstrap a "pivotal" statistic such as the t-value rather than bootstrapping directly non-pivotal statistics such as a regression parameter.

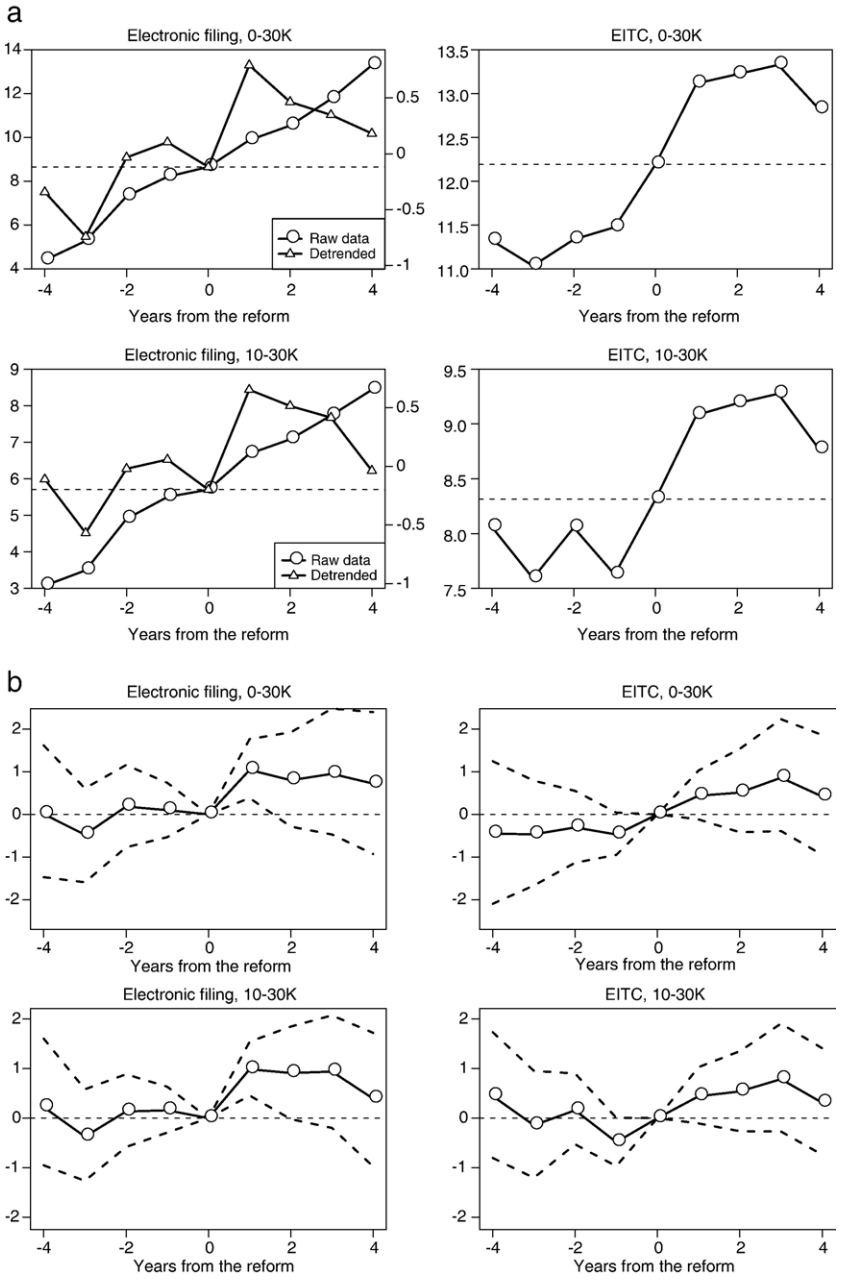

Fig. 3. Introduction of state electronic filing programs. (a) Raw data. *Notes*: The vertical axis represents the average share of electronics filers (raw data, left scale) or EITC recipients in the 0–30K category (top) or in the 10–30K category (bottom). The right scale on the electronic filling graphs represents the average deviation from the trend obtained by regressing the relevant variable on year fixed effects. The horizontal axis represents the distance (in years) from the year of introduction of the state electronic filling program. (b) Regression based. *Notes*: Figures show coefficients based on a version of Eq. (1) that includes both leads and lags and state-specific trends. They also show 90% confidence intervals (constructed using block bootstrap procedure). All states with income tax are used.

Table 2
Baseline regression, 10–30K category

	Shares			Shares, balanced		Logs	IV
	Unbalanced		All	4-years	3-years	Unbalanced	Shares
	Electronic filing						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
State EFI	0.716	1.049	0.983	0.626	1.125	0.108	
	(0.378)*	(0.421)**	(0.416)**	(0.404)	(0.494)**	(0.057)*	
Clustered	0.088	0.034	0.039	0.182	0.055	0.106	
Bootstrap	0.058	0.015	0.020	0.124	0.029	0.062	
	Earned Income Tax Credit						
State EFI	0.719	0.834	0.824	0.587	0.975	0.124	
	(0.252)***	(0.352)**	(0.348)**	(0.621)	(0.492)**	(0.050)**	
Electronics filers							0.795
							(0.037)***
Clustered	0.011	0.041	0.038	0.408	0.088	0.030	
Bootstrap	0.005	0.019	0.018	0.351	0.049	0.015	
N	504	504	612	312	372	504	504
States	42	42	51	26	31	42	42
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State trends	No	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The analysis is for the period 1988–1999. The dependent variables are measured in either shares or logs and refer to the 10–30K income category. All regressions include year and state fixed effects and in most cases include state specific time trends (as indicated). Standard errors are in parentheses and are based on the standard block bootstrap procedure. In addition, for the main variable of interest (State EFI), the table also presents *p*-values based on clustered standard errors (at the state level). ***, ** and * indicate statistical significance at the 1, 5 and 10% levels respectively.

specification of the type described by Eq. (1). In column (1) we show a specification based on an unbalanced panel that only includes year and state controls. Our preferred specification (column 2) in which we include state-specific time trends, shows that there is a large and statistically significant increase in the number of federal electronic filers following the introduction of such state programs. In column 3, we include states that do not have a state income tax and therefore no state electronic filing program. Inclusion of such states presumably helps in identifying year effects and the results are robust to this extension. Columns (4) and (5) are similar to (2) but they are restricted to states that have at least 3 or 4 years of pre- and post-treatment data.[17]

The share specifications indicate that in an average state, and depending on the specification used, an additional 0.6 to 1.1 percentage points of the population files electronically. The log specification, shown in column (6), predicts a 10% increase in e-filing following the introduction of the program. These effects are large and statistically significant, except for the balanced specification with 4 years of pre- and post-data where the number of states is reduced to only 26. The robustness of our results across specifications suggests that our findings are not driven by

[17] The balanced results in the working paper version (Kopczuk and Pop-Eleches, 2005) used states with at least three years of data after the treatment thereby including a few states who were treated very early. The difference in results is small.

Table 3
Robustness checks

	Electronic filing							
	0–30K	0–10K	10–20K	20–30K	30–100K	Waivers	State EITC	Telefiling
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
State EFI	0.949	−0.100	0.802	0.247	−0.032	1.049	1.050	0.938
	(0.518)*	(0.234)	(0.331)**	(0.172)	(0.167)	(0.421)**	(0.422)**	(0.397)**
State waiver						0.052		
						(0.333)		
State EITC							0.083	
							(0.809)	
State Telefiling								1.112
								(0.132)***
Clustered	0.113	0.708	0.037	0.215	0.865	0.034	0.035	0.044
Bootstrap	0.064	0.669	0.016	0.155	0.851	0.015	0.015	0.020
	Earned Income Tax Credit							
	0–30K	0–10K	10–20K	20–30K	No EITC, 0–30K	Waivers	State EITC	Telefiling
State EFI	0.716	−0.119	0.583	0.251	−0.468	0.837	0.832	0.828
	(0.430)*	(0.200)	(0.266)**	(0.151)*	(0.482)	(0.343)**	(0.352)**	(0.349)**
State waiver						0.386		
						(0.254)		
State EITC							−0.230	
							(0.352)	
State Telefiling								0.062
								(0.087)
Clustered	0.151	0.603	0.060	0.142	0.385	0.036	0.041	0.041
Bootstrap	0.095	0.559	0.029	0.096	0.328	0.017	0.020	0.019
N	504	504	504	504	504	504	504	504
States	42	42	42	42	42	42	42	42

Note: The analysis is for the period 1988–1999. The dependent variables are measured in shares. All regressions include year and state fixed effects and state specific time trends. State EFI refers to the average treatment effect. Standard errors are in parentheses and are based on the standard block bootstrap procedure. In addition, for the main variable of interest (State EFI), the table also presents *p*-values based on clustered standard errors (at the state level). ***, ** and * indicate statistical significance at the 1, 5 and 10% levels respectively.

compositional changes that might arise in an unbalanced panel. At the same time, leaving out state trends does not have a major effect on the estimated coefficients.[18]

These results provide clear evidence that the state electronic filing indeed had a significant impact on the number of federal returns that are filed electronically and thus we conclude that these changes provide an opportunity to identify the economic impact of electronic filing in general.

Table 3 presents the results for other income categories. The results are significant when we use the broader 0–30K income groups, as are all the specifications in the main table when using the 0–30K group (results not reported). Breaking the income groups into smaller categories

[18] However, the case for having state-specific trends is strong here: because the composition of population varies across states, penetration of electronic filing as measured by the share of population relying on it is very likely to systematically vary. Consequently, during a period of a rapid expansion of electronic filing, trends for different states are almost certain to have different slopes.

shows that the effect of electronic filing is significant only in the 10–20K category. The lack of response for the wealthier individuals is not particularly surprising because the impact of changes in electronic filing was unlikely to have played a significant role for them. The value of receiving a refund quickly – the main channel that we suggest in Section 5 – most likely falls with income and therefore it appears unlikely that many higher income individuals would be motivated to e-file by the introduction of state e-filing.

It is more surprising to observe that e-filing in neither the 0–10K category nor the 20–30K category appears responsive to the policy changes. Regardless of the reason (one problem is a discrete change in eligibility for the 0–10K category in 1994), it suggests that any response to treatment of our other variables of interest in these categories (such as EITC claims) should be treated with caution and could indicate that the relationships we estimate are spurious. However, there are possible explanations for a lack of response in e-filing in these categories that is consistent with observing effects in the other variables. As an example, it may be that increased e-filing in, say, the 0–10K category is coupled with behavioral responses that lead many individuals to shift to the 10–20K category. Such a behavioral response could in principle lead to e-filing being relatively stable, but it would simultaneously lead to a reduction in the number of individuals in the 0–10K category. We will return to this issue in the next section. Finally, the EITC-related motives for e-filing are strongest in the 10–20K category, because it includes the plateau region where benefits are maximized.

4.3. EITC claims

Next we turn to the impact on EITC claims. The second panel in Table 2 follows the same pattern as the panel for electronic filing. The results show a large positive effect of our treatment on the number of people claiming EITC on their federal returns. These are the main results of this paper. The results in Table 2 are only slightly sensitive to the inclusion of state-specific time trends.[19] The results using a "balanced" panel of states are very similar to the baseline ones.[20] The final specification shows the instrumental variable estimate of the effect of e-filing on EITC participation using the four treatment dummies as instruments. The instrumental variable estimate in the share specification has a straightforward interpretation: 79% of an increase in e-filing due to the introduction of e-filing programs corresponds to new EITC participants. In Table 3, we present the results by income categories. Consistently with the findings for electronic filing, we find no evidence of response in either the 0–10K category or the 20–30K category.

Since our identification strategy relies on the differential timing of the implementation of these programs, we also considered adding four leads and four lags to our specification in order to better understand if the shift in electronic filing and EITC claims happens around the time of program introduction. The coefficient estimates for these regressions are presented both for the full and the balanced panel in Fig. 3b, corresponding to specifications (2) and (5) in Table 2. The graphs are normalized relative to the zero coefficient on the omitted dummy for the year immediately preceding treatment (year zero); and they also include confidence intervals (at the 90% level) for these coefficients from the regressions. There is a clear pattern of the effect at time zero, and

[19] Contrary to e-filing, EITC claims are not strongly trending and therefore state differences in the rate of growth of EITC claims are not too important. Aggregate trends appear to control appropriately for any time effects.

[20] It should also be noted that in all specifications in this paper states are weighted equally. We found some evidence that the effect is stronger for smaller states.

although confidence intervals are generally large (not surprisingly, restricting instead lags to zero and leads to be equal to each other as in our tables increases precision of estimates), many of the post-treatment dummies are marginally significant at the 90% level. The number of electronic filers in the period prior to the implementation is constant and statistically indistinguishable from zero. Similar patterns are evident in the figures that plot EITC claims around program implementation. Despite relatively wide confidence intervals, the pattern of an increase in EITC claims following time zero is visible. These figures substantiate the main result of our analysis: the introduction of electronic state income tax filing programs resulted in more people claiming the Earned Income Tax Credit.

In Table 3 we present a few additional specification checks. Schoeni and Blank (2000) argue that state-specific welfare waivers passed in the early 1990s affected labor force participation and that the incremental impact of the 1996 reform was small. Therefore, we have re-estimated our specifications while controlling for the pre-1997 presence of a state waiver (column 6). Our results were virtually unaffected even though the state waiver dummy potentially had an effect on EITC participation (the coefficient is insignificant though), but not on e-filing. The introduction of waivers is plausibly a relevant consideration for studying EITC participation, but appears orthogonal to state e-filing changes. In the next specification check, we control for the presence of a state level EITC program and it appears to have no effect on federal EITC claims.

Finally, in column 8 we control for the possibility of TeleFiling which is an alternative to electronic filing by a tax preparer. Returns filed through TeleFile are included in our measure of e-filing, but TeleFiling could only be used for non-complicated returns and did not allow for claiming EITC. States were introducing their own TeleFiling programs throughout the 1990s, but usually much later than their regular e-filing programs. By 1995 only 3 states had the program in place and 11 states were yet to implement the program in 2001. Unsurprisingly, but very consistently with our story, we find that TeleFiling affects the number of returns filed electronically. It does not however affect the number of EITC claims — an expected finding given that TeleFiling cannot be used by taxpayers with dependents, who are the only ones eligible for EITC in the 10–30K income category.

In other robustness tests (not shown in the paper),[21] we dropped particular years or states from our sample and found the main results unaffected by these changes. In addition we accounted for possible changes in the number of tax filers by running the main specification using state population as the denominator.We also defined treatment solely on whether and when a state introduced the *Federal/State Electronic Filing Program*. In order to control for state specific welfare programs meeting TANF requirements introduced after September 1996, we re-estimated our baseline specification on the pre-1997 observations. We also allow for the possibility of a different regime pre-1994, when the IRS tightened its enforcement of EITC and post-1996 following the welfare reform, by introducing the interactions of state fixed effects with pre-1994 and post-1996 dummy respectively. The tightening of enforcement could have affected states where tax preparers were more prevalent more than other states and this could bias our estimated coefficients. Finally, we also included state labor force participation rates, the unemployment rate, the logarithm of employment and crime rates as dependent variables and found no evidence that they respond to the timing of state e-filing programs. This provides an added comfort that we are not picking up the effect of some omitted variable that affected state socioeconomic conditions in a way correlated with the timing of our treatment.

[21] These additional robustness tests are discussed in more detail in the working paper version (Kopczuk and Pop-Eleches, 2005).

5. Conclusion and discussion

Our analysis provided evidence that the introduction of electronic filing led to an increase in EITC participation. We demonstrated this link by showing that the increase in both federal electronic filing and EITC participation was much stronger in states that introduced the possibility of electronic filing of state income tax returns. In our analysis we provided a number of specification checks to test the robustness of the results and to rule out competing explanations, such as other welfare, EITC and IRS related reforms implemented during the 1990s.

We conclude with a discussion of the possible mechanisms that could be driving our results. Our analysis has not emphasized so far the role that the tax preparation industry has played in the adoption of electronic filing of tax returns. The introduction of electronic filing was from the very beginning a partnership between the IRS and the tax preparation industry. In fact, in 1985, the first pilot electronic filing program was a partnership between the IRS and H and R Block, the largest tax preparation company in the market.[22] The advantages to the tax industry of this partnership were twofold. First, at least until the late 1990s when on-line filing began to take off, anyone interested in filing taxes electronically basically had to use the services of the tax preparation industry. According to an IRS insider "when the program began, the only way to electronically file a 1040 was to pay a private tax preparer for the privilege" (Davis, 1998, p. 70). Secondly, electronic filing allowed the tax preparation industry to actively market Refund Anticipation Loans (RALs) that are one of its most important profit sources.[23,24]

The introduction of electronic filing as a new technology for tax preparation had a large effect on this particular industry. The founder of Jackson Hewitt was "mapping the company's expansion quite deliberately over the years to the IRS's test sites for its e-filing program [and] the plan helped make Jackson Hewitt, now a leader in providing electronic tax services, the second largest tax preparation company in the nation."[25] Similarly, the introduction of state electronic filing provided an additional incentive for tax preparers to provide e-filing services because it introduced the possibility of going all-electronic, reducing the paperwork and increasing the size of potential refund for the purpose of offering RALs. The importance of the joint state/federal electronic filing programs in inducing the tax preparation industry to adopt and popularize e-filing as a method for filing federal tax returns has been repeatedly highlighted by the IRS. Thus in a recent report to Congress, the Electronic Tax Administration Advisory Committee (2002) emphasized that "the ability to simultaneously file both a federal and a state return adds significantly to the value of e-filing for taxpayers and tax practitioners, and will lead to continued growth of e-filing at the federal level."

Finally, there is also some evidence that the expansion of this industry was targeted towards low-income neighborhoods. Based on the data for the largest 100 metropolitan areas, Berube et al. (2002) found that in 1999 zip codes with a share of EITC recipients greater than 30% had on average 15 electronic return originators per 10,000 taxpayers, while those with the share below 5% had on

[22] See Fletcher (2003).

[23] Some industry experts consider the introduction of RAL's as the chief incentive driving the volume of electronic filing. (Source: December 6, 1994 speech by Ms. Peggy Rule, IRS Director for Electronic Filing Programs, before the AICPA Tax Division, Retrieved May 20th, 2004 from the World Wide Web: http://www.nysscpa.org/prof library/Briefingbook/Spring96/tax2.tbb.htm.)

[24] RALs and similar quick refund products contributed 29.8% of total revenue for Jackson Hewitt, the second largest firm in the market (Tax Notes, Sept. 22, 2003, p. 1488). On the day of the 1994 announcement that the IRS will stop providing the Direct Deposit Indicator to e-filers, shares of H and R Block fell by 17% ("H and R Block Stock Plunges On Electronic Filing Change", New York Times, November 23, 1994).

[25] Source: Tax Notes, Apr. 9, 2001, p. 189.

average 10 electronic return originators per 10,000 taxpayers. They also estimate that almost half of 1999 EITC dollars involved RALs.

We speculate, that the expansion of electronic filing increased the incentive for tax preparers to market their services and go after customers who would not otherwise file a tax return. In particular, the marginal cost of providing a quick refund was reduced significantly due to a reduction of the risk associated with this product. Furthermore, the IRS was instrumental in making sure that the alternative means of obtaining at least one of the services provided by tax preparers (i.e., quick refund) were unavailable. Although our results are consistent with the view that there are potential benefits of the decentralized tax preparation industry, the use of RALs and the role of tax preparation industry remain controversial, because they reduce the actual benefits received by low-income individuals. At the same time, the increased uptake of EITC is also consistent with other possible explanations, such as the reduction in transaction costs for filers and increased information about the program. Separating the relative importance of the different mechanisms that could explain the main findings of this analysis is beyond the scope of this paper, and is left for future research.

References

Bertrand, Marianne, Duflo, Esther, Mullainathan, Sendhil, February 2004. How much should we trust differences-in-differences estimates? Quarterly Journal of Economics 119 (1), 249–275.

Berube, Alan, Kim, Anne, Forman, Benjamin, Burns, Megan, May 2002. The Price of Paying Taxes: How Tax Preparation and Refund Loans Erode the Benefits of the EITC," Survey Series, Center on Urban and Metropolitan Policy, The Brookings Institute and The Progressive Policy Institute.

Currie, Janet, May 2004. "The Take Up of Social Benefits," Working Paper 10488, National Bureau of Economic Research.

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

Electronic Tax Administration Advisory Committee, June 30 2002. "Annual Report to Congress," Publication 3415 (Rev. 6-2002). (Retrieved May 17th, 2004 from the World Wide Web: http://www.irs.gov/pub/irs-pdf/p3415.pdf).

Erard, Brian, Ho, Chih Chin, July 2001. Searching for ghosts: who are the nonfilers and how much tax do They Owe? Journal of Public Economics 81 (1), 25–50.

Fletcher, Patricia Diamond, 2003. "IRS E-File: Electronic Tax Administration in the United States," Case Study, Center for Technology in Government, University at Albany, SUNY. (Retrieved May 20th, 2004 from the World Wide Web: http://www.ctg.albany.edu/publications/online/newmodels/cases/irs.pdf).

Hill, Carolyn J., Hotz Joseph V., Mullin Charles H., Scholz John Karl, "EITC Eligibility, Participation and Compliance Rates for AFDC Households: Evidence from the California Caseload," April 1999. mimeo.

Holtzblatt, Janet, McCubbin Janet, "Complicated Lives: Tax Administrative Issues Affecting Low-Income Filers," mimeo, Department of the Treasury January 2003.

Horowitz, Joel L., 2000. The bootstrap. In: Heckman, J.J., Leamer, Edward (Eds.), Handbook of Econometrics, vol. 5. Elsevier/North Holland, Amsterdam, New York.

Hotz, Joseph V., Scholz, John Karl, 2003. The earned income tax credit. In: Moffitt, Robert A. (Ed.), Means-Tested Transfer Programs in the United States. University of Chicago Press.

Internal Revenue Service, January 31 2002. Participation in the Earned Income Tax Credit Program for Tax Year 1996," Research Project 12.26, U.S. Department of Treasury.

Kopczuk, Wojciech, Pop-Eleches, Cristian, November 2005. "Electronic Filing, Tax Preparers and Participation in the Earned Income Tax Credit," Working Paper 11768, National Bureau of Economic Research.

Liebman, Jeffrey B., December 2000. Who are the ineligible EITC recipients? National Tax Journal 53 (4), 1165–1185 (Part 2).

Maag, Elaine, February 2005. "Paying the Price? Low-Income Parents and the Use of Paid Tax Preparers," New Federalism: National Survey of America's Families B-64, Urban Institute.

McCubbin, Janet, December 2000. EITC noncompliance: the determinants of the misreporting of children. National Tax Journal 53 (4), 1135–1164 (Part 2).

Schoeni, Robert F., Blank, Rebecca M., March 2000. "What has Welfare Reform Accomplished? Impacts on Welfare Participation, Employment, Income, Poverty and Family Structure," Working Paper 7267, National Bureau of Economic Research.

Scholz, John Karl, March 1994. The earned income tax credit: participation, compliance and anti-poverty effectiveness. National Tax Journal 48 (1), 59–81.