Electronic Filing, Tax Preparers, and Participation in the Earned Income Tax Credit

Wojciech Kopczuk* and Cristian Pop-Eleches†‡

First draft: February 17, 2004
This draft: March 11, 2005

*Columbia University and NBER, e-mail: wkopczuk@nber.org
†Columbia University, e-mail: cp2124@columbia.edu
‡We are grateful to Michael Conlin, Gary Engelhardt, Brendan O'Flaherty, Jeff Liebman, Lucie Schmidt, Till von Wachter and seminar participants at Columbia, Syracuse, Harvard/MIT and Duke for helpful comments.

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. Most of the tax returns for EITC recipients are filed electronically by paid tax preparers who often charge significant fees for their services. 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 suggest that this effect is due to the impact that electronic filing opportunities had on the tax preparation industry, therefore providing an example of how a market-based approach can be effective in addressing the problem of program non-participation.

1 Introduction

Tax rules and details of welfare programs are difficult to understand for economists, and even more so for the average taxpayer. The classic "tagging" logic emphasized by Akerlof (1978) suggests that better targeting of benefits allows for reducing distortions. At the same time, precise targeting requires specifying precise eligibility criteria thereby increasing the complexity of rules applying to individuals. As a practical matter, increased complexity makes it difficult for intended recipients to learn about and benefit from government's programs; they are also bound to make mistakes and to be unaware of all of their responsibilities. Consequently, increased tagging does not just lead to reduced distortions but also limits the effectiveness of transfer programs by making them more complex. We provide empirical evidence, in the context of the Earned Income Tax Credit, that this trade-off is endogenous and sensitive to its institutional and technological environment. In particular, we argue that the *market* for tax preparation services effectively helps in addressing the problem of non-participation.

Complexity is often suggested as the reason for reforming the US tax code and it is also one of the key issues in designing welfare programs. In this paper we focus on a particular program, the Earned Income Tax Credit (EITC), which has attracted a lot of attention in recent years from both academic and policy circles. While we discuss the evidence and difficulty of measuring non-participation in this program in detail in Section 2.2, the main puzzle behind the present analysis is that non-participation in the EITC is non-trivial despite its significant advantage over other types of welfare programs. This advantage is generally attributed to its integration into the income tax system; therefore, individuals who file their taxes are likely to be aware of its existence.

In this paper we demonstrate the causal link between the technology of tax compliance and claiming of EITC benefits. By reducing the time it takes to obtain a refund, the introduction of electronic filing technology constituted a breakthrough in the tax preparation industry. This change provided tax preparers with the opportunity to provide a new kind of service: a quick refund. Filing taxes electronically usually required (at least at the lower end of income distribution) relying on a tax preparer. As the result, the industry was able to reap benefits of this new technology. At the same time, the additional rents made it worthwhile to expand and reach taxpayers who would otherwise not file their tax returns. Thus, our results provide support for and are an example of the possibility discussed by Currie (2004) in a recent review of the literature on the take up of social programs, that participation in such programs can be increased when businesses or individuals have a stake in promoting them.

More specifically, we use the staggered introduction of state electronic filing programs during the 1990's 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. To test the robustness of our main result, 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, IRS programs to reduce EITC related cheating, or the reform and expansion of EITC in 1991 and 1994. We provide evidence that this growth cannot be explained by cheating of existing filers and show that the bulk of the effect represents increased participation by individuals who previously did not file their tax returns ("non-filers"). While we provide some evidence that these new filers of tax returns and EITC benefits are similar among many observable characteristics, we cannot rule out entirely the possibility that cheaters might drive some of the increase in participation.

The results suggest a counter-balancing argument to claims that the tax preparation industry takes advantage of EITC claimants. Tax preparers charge significant fees for their services. Although electronic filing itself very significantly reduces the time from filing to the refund, tax preparers often provide the so-called Refund Anticipation Loans (RALs) to electronic filers. These loans have benefits to taxpayers: they amount to receiving their refund earlier (in some cases, instantaneously) and may relax liquidity constraints. However, these loans involve little risk[1] and given the relatively short time that it takes for the IRS to process an electronic return, these fees amount in some cases to exorbitant interest rates.[2] We do not evaluate the direct welfare benefits and costs of the RALs,[3] but our results suggest that one should not consider RALs out of context: in their presence tax preparers have an additional incentive to assure that taxpayers apply for the benefits they are due. As a result, the RALs may be considered to be the price of the (arguably socially beneficial) market solution to the problem of complexity.[4]

Our analysis proceeds as follows. We describe the EITC program, changes in tax filing technology, and the tax preparation industry in Section 2. In Section 3, we present a very

[1] In fact, tax preparers are even able to verify before they provide a loan whether the taxpayer owes taxes to the IRS in which case credit would be reduced.

[2] On the high cost of tax preparation services and RAL loans see Berube et al. (2002) and Wu and Fox (2003). Berube et al. (2002, page 14) are explicit about the negative welfare impact of the tax preparation industry: "There is no question, however, that the public interest is not served when a tax credit designed specifically to supplement the earnings of low-income workers with children becomes an important profit center for a multi-billion dollar industry."

[3] A baseline economic model suggests that taxpayers who decide to purchase a loan of this kind must find it beneficial. In practice, this is more controversial because the tax preparation industry was accused of misleading the customers by blurring the distinction between rapid refunds and loans (see "H&R Block Is Sued by City Over Loan Ads," New York Times, March 31, 2002.) Implicit in this argument is that taxpayers have to be unaware of the possibility of receiving a refund quickly (though not as fast as in the case of RALs) by filing electronically without purchasing a loan.

[4] We of course do not claim that RALs are the optimal solution. The question of how and if the government should encourage market solutions intended to increase program participation is an interesting topic for future research.

simple model that illustrates how the introduction of electronic filing could have affected EITC participation. In Section 4, we introduce our data and empirical specification. The main empirical analysis is presented in section 5. We discuss the possible explanations and interpretations of our results in Section 6. We conclude in Section 7 with the discussion of implications of our results and possible extensions.

2 Earned Income Tax Credit and E-Filing

2.1 EITC

We will only briefly review the main features of the EITC, since Hotz and Scholz (2003) provide a recent survey on its history and research on this topic. The Earned Income Tax Credit was introduced in 1975 to provide support for the working poor. The benefits are phased in for low income workers, held constant in the middle range, and phased out for the highest eligible incomes. The EITC was expanded between 1991 and 1994. The 1990 EITC expansion was phased in over three years so that in 1993 the maximum benefit was higher than the 1990 maximum by almost 50% (in real terms). The 1993 budget bill further expanded the reach of the program by another round of benefits increases, and for the first time, expanded eligibility to childless individuals. EITC benefits vary with the number of qualifying children. The 1999 (the last year we use in our analysis) maximum for workers with two or more children was $3,816, with the phase in range between $0 and $9,540 and the phase out range beginning at $12,460. The credit was fully phased out at $30,580. Since 1994, single individuals are also eligible, however, in 1999 their maximum benefit was just $347 and the program was fully phased out when income reached $10,200. The Earned Income Tax Credit is refundable: if the credit exceeds taxpayers' tax liability, the excess is payable to the taxpayer.[5]

In order to claim the credit a taxpayer has to file a tax return. Taxpayers generally are not required to file if their income is below the filing threshold (the sum of standard deductions and own personal exemptions). If there are dependents, exemptions for them are not part of the filing thresholds so that the taxpayer may be required to file even if there is no income tax liability. Furthermore, there is also a significant group of individuals who do not file even though they are required to do so (see, Erard and Ho, 2001, for the discussion).

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. Only own children, grandchildren or foster children qualify. A child must be either under 19 (24 for a full-time student) or be permanently and totally disabled. The

[5]In 2002, eighteen states have 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.

residency test requires that the child lives with the taxpayer for at least half a year in the case of an own child and the whole year in the case of a foster child. If two different taxpayers are eligible, the one with the higher AGI is supposed to claim the child. The qualifying child criteria for EITC are different from those for child tax credit and tax exemptions. In fact, it is possible that a taxpayer can claim tax exemption for the child that's not a qualifying dependent for the purpose of EITC and vice versa. Apart from the criteria related to children, there is also room for manipulating the system by adjusting filing status. While a married couple filing jointly is eligible for the credit, the resulting credit is smaller than if they each filed separately. Another source of potential fraud and complexity within the system has to do with the definition of income for EITC. Generally, public assistance benefits are supposed to be excluded from calculation of income for the EITC purposes. Certain types of income can make taxpayer ineligible if they exceed the threshold amount ($2,350 in 1999). They include interest, dividend, rent and royalties, capital gains and passive income not related to self-employment.

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). During audits agents were also asked to indicate whether they believed that mistakes were intentional and based on this information the IRS officials testified that approximately 50 percent 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. Most of the studies on EITC participation rely on datasets that allow to determine eligibility, such as the Census Bureau's Current Population Survey March Supplement (CPS) or the Survey of Income and Program Participation (SIPP). However, even studies based on confidential administrative data sources such as Internal Revenue Service (2002) are unable to determine with certainty whether an individual claimed EITC and instead rely on proxies such as whether the tax return was filed.

Scholz (1994) estimated that in 1990 approximately 80% of eligibles actually claimed benefits. A recent IRS study (Internal Revenue Service, 2002) states that in 1996 between 13% (based on march 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. However, as discussed below, 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%.

Secondly, Hotz and Scholz (2003) speculate that non-participation in the 1990's 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.

Many advocacy groups concentrate their effort on providing information about the program[6]. The IRS itself continues to devote a lot of attention to the issue of non-participation. In a recent press release, the IRS listed a number of ways in which it attempts to increase participation including "working with more than 180 community-based organizations nationwide to reach low-income workers who may be unaware of the EITC availability", "helping set up 14,000 volunteer centers that offer free tax preparation for low-income and elderly individuals", "coordinating with mayors offices nationwide to help identify low-wage earners who may qualify for EITC" and "teaming with a special grassroots cadre in two target cities — Los Angeles and Miami —to reach out to taxpayers who have limited proficiency in English but who may qualify for EITC."[7]

2.3 E-filing

The IRS introduced electronic filing of tax returns in 1986. Electronic filing is beneficial from the IRS point of view because it reduces the cost of handling paper returns and the

[6]For example, one of the websites devoted to the EITC (`http://www.eitc.info/eitc_basics/`, accessed on December 3 rd, 2003) states as its objective "Ensuring that more families know about and file for the tax credits they are due."

[7]`http://www.irs.gov/newsroom/article/0,,id=119792,00.html`.

likelihood of mistakes.[8] Taxpayers can benefit from electronic filing in a number of ways. First, returns filed electronically have a much lower error rate than paper returns therefore reducing the risk of audits and penalties. Second, the more efficient handling of the returns by the IRS and the reduced likelihood of mistakes both facilitate a much quicker potential refund. Third, since software has to be involved in order to file a tax return electronically, it increases the chance of discovering previously unknown opportunities for tax saving. Fourth, taxpayers receive an immediate confirmation that the return was accepted.

There are a few different ways to file tax returns electronically, and patterns of filing changed over time. 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. Throughout the paper we will be focused on electronic filing; whether a professional signed the tax return is secondary. In the late 1990s, many taxpayers chose to file their tax returns by phone under the TeleFiling program. This option has always been limited to taxpayers with relatively simple tax returns: the reach of this method of e-filing is limited to the group of childless taxpayers. These taxpayers have low income maxima for EITC and low maximum benefits. Because we are not able to distinguish between telefiling and other forms of e-filing in our data, we consider this low-income group separately in our analysis. Finally, the direct on-line filing is another possibility to file electronically, but its importance was in the period that we consider limited.[9]

Forty one states and the District of Columbia operate their own income tax systems. The relationship of state and federal income taxes varies, but in each case they are operated as separate systems. Traditionally, taxpayers were (and still are) required to file both federal and state tax return. States started introducing their own electronic filing systems in the early 1990s, a few years after federal electronic filing was introduced. 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. 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. 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. The remaining four states operate their own electronic systems. Two states (Illinois and Maryland), operated their own electronic filing systems before they joined the federal program. The timing of introduction of state electronic filing varied between 1990 for South Carolina and 2001 for Hawaii and Vermont. We later show that the introduction of state electronic filing had significant effect on the number of federal

[8] According to the IRS "Questions and Answers" about e-filing using a tax preparer (`http://www.irs.gov/efile/page/0,,id=10093,00.html`), accessed February 15, 2004), the error rate for electronically filed returns is less than 1%, compared to the error rate for paper returns which is about 20%.

[9] 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.

returns that were filed electronically.

The FedState program also produced other types of interactions between federal and state tax authorities including sharing computer files with taxpayer information, sharing audit results and organizing seminars for tax practitioners, cooperation in administration of other types of taxes. By the late 1990s, the program also led to the introduction of joint state/federal TeleFiling systems that allowed the qualified taxpayers to file their taxes over the phone. While the federal TeleFiling system was tested in Ohio as early as in 1992 and became available nationwide in 1996, only 3 states introduced it before 1996; 11 still do not have one in place. The joint state/federal TeleFiling possibility was only introduced for the returns filed for 1998 in Indiana and Kentucky, and by 2003 it was still only available in seven states. Although, as discussed above, claiming EITC using TeleFile is not a concern, it is possible that media/marketing coverage of this new possibility might have attracted attention to other types of electronic filing possibilities. In particular, we will show as a robustness check that the number of e-filers responds to state TeleFiling programs but EITC participation does not.

2.4 The Role of Tax Preparers

We use the term "tax preparer" to describe any professional involved in the process of filing a tax return, including electronic return originators (so-called EROs) who file electronically tax returns prepared by their customers. In terms of the structure of the tax preparation industry, two large companies dominate in the market: H&R Block and Jackson Hewitt. Together they controlled roughly 30% of the electronic filing market in 2001 (15.6% of all returns filed). The remaining 70% is highly diversified and includes both trained professionals and amateurs who often work only during the tax season.

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&R Block, the largest tax preparation company in the market.[10] The advantages to the tax industry of this partnership were twofold. First, at least until the late 1990's 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, the IRS was willing to share information about filers with the tax preparers. Tax preparers actively market Refund Anticipation Loans (RALs). They are very short term loans made usually by a tax preparer as the intermediary of a bank to an electronic taxpayer once a tax return is filed. RALs are due to be repaid when the refund is received from the IRS. For every tax return submitted electronically, the IRS would immediately issue a direct

[10] See Fletcher (2003).

deposit indicator (DDI), also known as the debt indicator or the refund offset indicator. This indicator informs the tax preparer if the refund requested by the taxpayer will be reduced because of a debt owed to the IRS or other federal agencies, such as outstanding child support or student loans. Therefore, the existence of this debt indicator greatly reduced the risks associated with RALs. Some industry experts consider the introduction of RAL's as the chief incentive driving the volume of electronic filing.[11] The IRS further reduced the risk of RALs by making it possible that the refund be deposited directly to the creditor's account.

As early as 1990, it was possible to file taxes electronically at many locations across the country. For example, a company called Instatax had as many as 7000 sites in 1990 located in all states at major retail chains such as Mailboxes Etc., 7-Eleven, or Revco drugstores.[12] At these particular locations, no actual tax preparation was done; instead customers brought returns prepared by themselves or by outside tax professionals. These returns were faxed to the company headquarters, verified for inconsistencies, and filed electronically with the IRS. Refund Anticipation Loans were provided through an arrangement with the Greenwood Trust Company once the IRS confirmed that the refund was forthcoming. The fee for electronic filing was $29.95 and the fee for the loan was $35. Fees for the RAL in the following years appeared higher. A newspaper article in 1992[13] reported fees at H&R Block to be $63. Some places in the Atlanta area charged as much as $112. By 1995, the fee for a $1,000 loan was reportedly $89, even though the loan was typically repaid within 13 days.[14] The reported fees for RALs are probably not a reliable measure of their actual cost, because the RALs are purchased as a part of a package with electronic filing and may also involve tax preparation. There is usually also a processing fee involved. According to a 2003 report by a consumer advocacy group (Wu and Fox, 2003), a tax filer could expect to pay on average $75 for a RAL fee, $40 for an electronic filing fee, $100 for a tax preparation fee and another $33 for additional processing fees. The estimated drain of all these fees on the overall EITC program was around $1.2 billion, roughly 4% of total EITC spending.[15]

Berube et al. (2002) document that commercial tax preparers target low income neighborhoods. Based on the data for the largest 100 metropolitan areas, they found that in

[11] 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`.

[12] This paragraph is based on Jan M. Rosen, "Electronic Filing For Tax Returns", New York Times, February 10, 1990.

[13] "Impatience proves costly with 'instant tax refunds' Fees of more than $100 aren't unusual", Atlanta Constitution, February 7, 1994

[14] David Cay Johnston, "IRS and Lender Settle Dispute on Refunds," New York Times, March 4, 1995.

[15] Note that even if we add the administrative costs to the IRS (estimated at around 1% of EITC benefits) to the above mentioned costs related to the preparation of returns, the total costs of EITC still compare very favorably to the administrative costs of, for example, AFDC (Liebman, 1998, estimated in 1995 at 16% of benefits paid). Even if we only use as a base of this calculation the 50% of EITC claimants who e-file, we are left with a number smaller than 10% of the total spending on the program.

1999 zip codes with a share of EITC recipients greater than 30% had on average 15 EROs per 10,000 taxpayers, while those with the share below 5% had on average 10 EROs per 10,000 taxpayers. They also estimate that almost half of 1999 EITC dollars involved RALs.

The importance of IRS cooperation in securing Refund Anticipation Loans was stressed by the industry. When the IRS eliminated the Direct Deposit Indicator and announced that it will send refunds directly to taxpayers in November of 1994,[16] some experts in the industry predicted that delinquency rate on RAL will increase from the extreme low of 1% to as much as 20% to 40%.[17] There is evidence that electronic filing and Refund Anticipation Loans played an important role for tax preparation companies. On the day of the 1994 announcement that the IRS will stop providing the Direct Deposit Indicator to e-filers, shares of H&R Block fell by 17%.[18] Similarly, RALs and similar quick refund products contributed 29.8% of total revenue for Jackson Hewitt, the second largest firm in the market.[19]

The introduction of electronic filing as a new technology for tax preparation had a large effect on this particular industry. Thus, 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."[20]

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. Benefits from electronic filing, such as speedy refunds, are also naturally increased from the point of view of the taxpayer when two rather than one returns can be filed in this way.[21] 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".

The significant role that electronic filing of state taxes plays for the tax preparation

[16] This policy was in effect during the 1995 season (i.e., 1994 tax returns) only. We test the robustness of our results to this policy change in the empirical section.

[17] David Cay Johnston, "Bank Challenges I.R.S. on Refunds for Borrowers", New York Times, February 22, 1995.

[18] "H&R Block Stock Plunges On Electronic Filing Change", New York Times, November 23, 1994.

[19] Source: Tax Notes, Sept. 22, 2003, p. 1488.

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

[21] Another channel through which state electronic filing might have stimulated federal filing is through advertisement campaigns. For example, the Wisconsin Department of Revenue actively advertised the program to the tax preparers (see "Wisconsin: DOR Encourages Electronic Filing," August 23, 1993, 5 State Tax Notes 403.

industry can be seen from the recent development of a free e-filing program of federal taxes for low income people, a cooperation between the IRS and the Free File Alliance, an organization of the tax preparation industry. Maybe not surprisingly, almost all participants in the free federal e-file program charge a fee for e-filing of state taxes, equal in the case of H&R to $24.95 for the 2003 filing season.[22] Similarly, according to a representative from a leading provider of electronic filing software for the tax preparation industry, the ability of their product to handle state electronic filing allowed their "customers to get an early start on their tax season and a jump on their competition."[23]

3 A Model of the Effect of E-Filing on EITC Participation

In order to highlight how the introduction of electronic filing can affect EITC participation, we present a simple model. Assume that there are many (heterogeneous) communities. There are the same number of individuals in each community, normalized to equal one. In community i, α_i individuals are eligible for EITC and $1-\alpha_i$ are not. The distribution of communities is characterized by the c.d.f. $F(\alpha)$ and the total number of communities is normalized to one. One straightforward interpretation of a community is as a geographical unit, but it can also be interpreted as a particular customer base. The key assumption of the model will be that it is costly for a tax preparer to reach a "community." The model applies both to the situation where a tax preparer makes a decision of whether to open a new location and to a situation where a tax preparer decides to serve a wider customer base from an existing location.

Preferences of individual j are characterized by $c_j - L_j$ where c_j is the consumption level and L_j measures the importance of liquidity constraints and therefore the value of receiving the refund quickly. Fraction γ of eligible individuals are liquidity constrained, so that $L_j = L$ (where L is a constant) when they have to wait for a refund and $L_j = 0$ otherwise. The remaining individuals are not liquidity constrained, so that $L_j = 0$ regardless of the timing of the refund. Consequently, the first type of individuals will decide to select a speedy refund if its price is lower than L while others will never buy it.

The eligibles are not necessarily aware of the EITC. We assume that either everyone in the community is aware of the program or nobody is. In order to stress the role of tax preparers, we assume that a community is aware of the EITC if and only if a tax preparer decides to locate there. A tax preparer decides to enter the community if it is profitable to do so. Entry requires paying the fixed cost of f. Although in general tax preparers provide many layers of service including preparation of the return, electronic filing and refund anticipation loans, we simplify by assuming that the product that tax preparers sell

[22] H&R Block advertisement, `http://www.hrblock.com/taxes/partner/product.jsp?productId=54`, retrieved from the World Wide Web May 18th, 2004.

[23] Drake Software, source: `www.taxingsubjects.com/archives/issue11 articles/artcl1.html`, retrieved from the World Wide Web on May 18th, 2004

is a speedy refund. In practice, both electronic filing alone and e-filing coupled with RAL fall into this category: both of these accelerate the refund — in the latter case the refund is received immediately, in the former it is received significantly faster than if the return was filed in the traditional way.[24,25] Critically, we assume that following the entry taxpayers who become aware of EITC and who value speedy refunds are not able to use alternative means of obtaining the refund. As discussed earlier, it appears to be a fair description of the situation throughout the 1990s, because electronic filing not through a commercial tax preparer was next to impossible for anyone but fairly computer savvy individuals.

The marginal cost of providing a speedy refund in period t is given by d_t. Once the tax preparer decides to enter the community, the maximum price that can be charged for a speedy refund is L — the value of the refund to liquidity constrained taxpayers. When the price is lower than L, all liquidity constrained taxpayers will decide to purchase the refund. Given the price $p_i \leq L$ of the speedy refund, the tax preparers' profit in community i is equal to $\pi_t(p_i) = (p_i - d_t)\alpha_i\gamma - f$. There will be entry to the community whenever profits at the maximum feasible price of L are positive:

$$\pi_t(L) \geq 0 \qquad \Longrightarrow \qquad \alpha_i \geq \frac{f}{\gamma(L - d_t)} \equiv \omega(d_t) \tag{1}$$

Given the cost of d_t, there will be entry and therefore the community will be aware of the EITC for sufficiently large values of α_i, i.e. the ones that exceed $\omega(d_t)$. Clearly, $\omega'(d_t) > 0$. Therefore, when the cost of a speedy refund goes down, the threshold for entry goes down as well and tax preparers enter additional communities.

Proposition 1. *Suppose that $d_{t+1} < d_t$. Then EITC participation in year $t+1$ is higher by $\beta(d_{t+1}, d_t)$ and the number of EITC participants who rely on tax preparer service increases by $\gamma\beta(d_{t+1}, d_t)$, where $\beta(d_1, d_2) = \int_{\omega(d_2)}^{\omega(d_1)} dF(\alpha)$.*

Proof. Following the decrease in the cost of a speedy refund, tax preparers enter communities with $\alpha \in \Big[\omega(d_{t+1}), \omega(d_t)\Big)$. The number of individuals in these communities is given by $\beta(d_{t+1}, d_t)$. Fraction γ of these individuals will purchase speedy refunds from tax preparers. □

The introduction of electronic filing corresponds to a reduction of the price of a refund. As discussed above, filing a tax return electronically significantly reduces the time it takes to

[24] In fact, using a tax preparer without e-filing may also be interpreted in a similar way: the reduced likelihood of a mistake amounts to a higher probability that a refund will be received without a delay.

[25] On its web page, H&R Block highlights the speed of receiving the refund as the critical factor in the development of e-filing: "In a 1986 test with the IRS, H&R Block filed 22,000 returns electronically from two sites. The test was a success: Electronic filing significantly reduced the amount of time required for a taxpayer to receive a refund. More than a decade later, H&R Block's trademarked "Rapid Refund" service has become synonymous with electronic filing. Now, the company files nearly half of the total number of returns filed electronically with the IRS." (`http://www.hrblock.com/presscenter/about/history.jsp`, accessed March 1st, 2004.)

receive the refund both due to faster processing by the IRS and due to the reduced likelihood of mistakes. Second, providing a loan toward a future refund became much less risky due to the IRS providing a debt indicator and making it possible to deposit the refund directly to the creditor's bank account, and the development of cooperative agreements among banks that facilitate identifying cheaters. The cost is further reduced when it is possible to file both state and federal tax electronically. The model predicts, therefore, that the introduction of electronic filing will increase the number of EITC recipients. However, only fraction γ of them will rely on tax preparers. The activity of tax preparers has a spillover effect of informing the remaining $1-\gamma$ of population about the EITC.

The model highlights two important aspects of the institutional environment that are important for increased participation in the program and these facts fit well with the actual developments in the United States in the 1990s. First, tax preparers have to be able to provide a product that is valued by low-income individuals. In the context of EITC, this product is the quick refund. As discussed earlier, the marginal cost of providing a quick refund was reduced significantly due to the introduction of electronic filing of tax returns. Second, because information about the existence of the program is non-excludable, the alternative means of obtaining the service provided by tax preparers (i.e., quick refund) must be unavailable or at least costly. By making it difficult to file electronically other than through a tax preparer, the IRS guaranteed that this condition was met. It is the combination of these two assumptions that leads to an increase in participation, as demonstrated by Proposition 1.

4 Data and empirical specification

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.[26] These samples are available for most years between 1960 and 1999 (with the exception of 1961, 1963 and 1965). 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 is 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.[27] The public use files contain most of the information from federal tax returns, but no other information.[28] For example, we can observe 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

[26]The documentation for this dataset can be accessed at `www.nber.org/taxsim`.

[27]In this dataset high income individuals are oversampled.

[28]Some information (including state of residence) is blurred or erased for high-income (above $200 thousand) taxpayers to protect taxpayer confidentiality. Since we are interested in the low income group only, this is not a major problem.

income, filing status and the number of dependents. We use sampling weights provided by the SOI.

We aggregate all information by state and income categories described below rather than using individual level data. 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 fact, previewing our results, we find that the most important channel of increased participation is through an increase in the number of tax filers: some individuals who used not to file a tax return started to do so after electronic filing was introduced. Furthermore, the existence of an electronic filing program varies only by year and state so that our source of identification is not compromised by aggregation. The standard argument for using individual level data is to control for observable individual characteristics. As we have already mentioned, our dataset contains very limited background information about filers. Moreover, of the variables included in the data, it is difficult to think of an observable personal characteristic that would not be endogenous: even demographic characteristics such as marital (filing) status or the number of children *as reported on the tax return* are likely to be responsive to EITC considerations because they affect eligibility and generosity of benefits. Furthermore, the response on the filing/non-filing status makes the distribution of these characteristics endogenous even in the absence of taxpayers' attempts to game the system.

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. The Tax Model Data is designed to be a representative sample of tax filers. We compared estimates of the number of EITC recipients as well as the number electronic filers by tax preparation status to the totals published by the IRS and they closely match.

We divided taxpayers into income categories based on their AGI.[29] Both adjusted gross income and earned income have to be below the relevant threshold (depending on the number of children and filing status) to be eligible for the EITC. All our dollar figures are expressed in 2000 dollars and adjusted for inflation. The maximum threshold between 1988 and 1999 was \$31,794 (in real 2000 dollars). The maximum threshold for childless individuals was \$10,597. Because the credit for childless individuals was introduced in the middle of our sample, in 1994, our major analysis is performed on the \$10,597-\$31,794 category that excludes childless EITC recipients. We refer to this category as "10-30K". We also present some results based on the \$0-\$10,597 category (and call it the 0-10K category) and divide the 10-30K category into two brackets by splitting it at \$20,000, thus creating

[29]The dataset contains a measure of the 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.

two additional income categories called the "10-20K" and "20-30K" categories. Finally we look at the whole category of individuals below $31,794 ("0-30K"). All EITC recipients in our dataset are included in this last category.

The major reason for this approach is to better account for changes in the criteria for receiving the EITC. The upper bound of earned income for receiving EITC benefits varied between $26,892 and $31,794 of real 2000 dollars. Since 1994, childless individuals are also eligible for receiving EITC benefits. The maximum income for receiving these types of benefits varied between $10,464 and $10,596. Because our dataset does not contain information about the number of children that are qualifying for EITC purposes, we are not able to precisely distinguish between different categories of EITC filers.[30] Changes in the number of low income EITC recipients are likely to be to a large extent driven by altered eligibility criteria. Dividing the sample into income groups allows us to isolate the impact of these changes. This approach will also make it possible to study potential shifting between income categories.

4.1 Summary Statistics

The first four columns of Table 1a 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 10% to 16% (an increase from 11 to 19 million). The table also highlights the major federal expansions of the EITC program in 1993 and shows that its effect is in fact limited to the 0-10K category.[31] The following four columns of Table 1a present analogous statistics for electronic filers. Although introduced in 1986, this federal program grew slowly in the first years after implementation so that by 1990 less than 10% of tax returns were filed electronically. Federal electronic filing grew steadily in the 1990's. One notable but complicating feature of the trend in electronic filing is the temporary fall between 1993 and 1994. Although the program was expanded at the same time, the source of the drop is due to attempts by the IRS to reduce cheating.[32] The measures involved both administrative requirements, such as the requirement to include the qualifying child's Social Security number on the return, as well as reduced cooperation with electronic returns originators. This change stopped

[30] We have attempted to classify taxpayers into different categories based on the size of credit. We were not always able to reconcile the amount of the credit with taxpayers income under any of the possible EITC claiming categories, although we were able to determine it for most of the taxpayers. We will show some results based on this classification as our robustness checks.

[31] 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 that does not include childless individuals, the number of EITC returns increased from 10.0 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.

[32] Table 1a demonstrates that this drop is most pronounced in the 10-30K income category that does not include newly eligible recipients.

the growth of e-filing for one year only. Furthermore, this change was likely perceived as temporary because the IRS had promised to facilitate usage of RALs in the future even before the end the same tax season. [33] Because this event falls exactly in the middle of our sample, we are not able to concentrate on years solely before or solely after it, but we will carefully investigate to make sure that this event is not behind our results.[34]

Trends in electronic filing by income category over this period are particularly interesting. For income groups in 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. 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. Figure 1 shows the overall trends in the major variables.

The last four columns of Table 1a summarize 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. However, the lower income groups experienced the largest uptake in relying on tax professionals and this increase is likely the result of the expansions of the EITC program during this period.

The extent of reliance on professional tax services among EITC claimants is described in Table 1b. Among EITC recipients, the proportion of EITC claims submitted electronically is large, growing over time and averages about 30%. The proportion of EITC recipients who do not prepare their own tax returns is also increasing and exceeding 60% by the end of the period. These trends for the 10-30K category are depicted in Figure 2. 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.

4.2 Empirical Strategy

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.[35] We

[33] David Cay Johnston, "IRS and Lender Settle Dispute on Refunds," New York Times, March 4, 1995.

[34] One might also note that there is no similar effect for EITC claims. Table 1a shows that there was virtually no change in the EITC claims between 1993 and 1994 for the 10-30K category.

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

verified this information (and corrected when necessary) against the IRS press releases.[36] Of the fifty states and the District of Columbia, nine do not have a state income tax: Alaska, Florida, Nevada, New Hampshire, South Dakota, Tennessee, Texas, Washington and Wyoming. In addition, two states North Dakota and Ohio introduced electronic filing in 1999, the last year for which the data are available, whereas Hawaii and Vermont only adopted this procedure in 2001. Of the remaining 38 states, six states (California, Illinois, Maine, Maryland, Massachusetts and Minnesota) first introduced their own independent program for electronic submission of state taxes but the electronic submission was not done jointly with the federal returns. The remaining states introduced joint electronic filing of federal and state income tax returns, under the IRS supervised Federal/State Electronic Filing Program. Of the six "independent" states Illinois and Maryland joined the StateFed program in 1995 while the four other states continue their separate programs.

Table 1c 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.[37] 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.[38]

Our identification strategy uses the variation resulting from the staggered timing of the introduction of electronic filing of state taxes as a way to estimate the effects of these programs on electronic filing of federal income taxes and its impact on federal EITC participation. Furthermore, we are also interested in understanding to what extent any potential increases in EITC uptake are explained by noncompliance. 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. Figure 3 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

[36] Available from `www.unclefed.com`

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

[38] 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.

large variation in the non-participation rates suggests that the pre-existing differences are not likely to explain the timing of the programs. Similarly, we plot in Figure 4 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.

One way to test the identifying assumption is to study if the timing of the state programs is related to any trends in our variables of interest. To preview our results, we show in Figures 5a, 5b, 6a, 6b that once controls are included in the regression framework, the timing of state electronic filing programs is not related in a systematic way to trends in electronic filing of federal returns prior to implementation or to trends in EITC participation. This would be the case in a situation in which state electronic filing is implemented only after federal electronic filing was increasing for other reasons.

We estimate OLS regressions of the following form:

$$outcome_{st} = \beta_0 + \sum_i \alpha_i statefiling^i_{st} + \beta_1 \delta_s + \beta_2 \tau_t + \beta_3 t\delta_s + \epsilon_{st}\,, \tag{2}$$

where $outcome_{st}$ is one of our dependent variables of interest (usually, electronic filers or EITC claimants), measured in logs or shares for each year and state cell in a given income category. The variables $statefiling^i$ are a set of dummy variables equal to one if a state had adopted state electronic filing i years ago[39] 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. We often restrict our specifications to include only the treatment dummy indicating whether a given state has an electronic filing program for state income taxes (it corresponds to the assumption that $\alpha_i = \alpha_j$ for any $i, j > 0$).

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 unbalanced 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 $statefiling^i$ 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 35 states that have at least 3 years of post treatment data. Finally we provide a number of additional robustness checks: we define treatment based on the timing of the joint federal/state filing programs rather than own independent electronic filing system, we code states without a state income tax as "always treated", we change the income brackets of interest, and we also ran our main specifications omitting individual states and years.

[39] We restrict i not bigger than 4 and define the fourth dummy variable to equal one if the state was treated four or more years ago.

In order to check the robustness of our results, 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. We do that 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. Furthermore, this approach allows us to potentially check whether cheating is an important component of the effect we identify. The welfare reform might have increased labor force participation (see Blank, 2002, for a review of the literature) and therefore the pool of potential EITC recipients. Because reforms were state-specific, it is possible that these effects are not accounted for by the inclusion of year fixed effects. If the parameters of state welfare reforms and therefore labor participation response were coincidentally correlated with the introduction of electronic filing programs, this could lead to a spurious relationship of EITC participation and the timing of state electronic filing programs.

We define our outcome variables in two possible ways: shares or logs. Shares 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. This approach will make it straightforward to analyze flows between income categories.

Defining our variables in logs is complicated by the fact that in the beginning of the period (1988-1990) we have a small number of zeros in the sample.[40] 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.[41]

Throughout the remaining sections, we present results for both shares and logs. Depending on the question, the interpretation of one measure may be easier than the other. In the case of the log specification, we either implicitly assume that the percentage effect of treatment is uniform across states or alternatively we evaluate the average percentage treatment. In the case of share specification, we make the corresponding assumption in

[40]The exact breakdown of the number of zeros in a given year is as follows: 20 states in 1988, 2 in 1989, 1 in 1990 and 1991 and no zero values afterwards. As discussed later in the text, results are robust to dropping zeros or starting in 1991.

[41]Changes in the size of this adjustment within a reasonable range do not affect our results in an important way. Also, as discussed later, dropping early years (when zeros occur) from the sample does not affect the results.

terms of the share of state's taxpayers. An added advantage of presenting the results in both shares and logs is that we are providing a robustness check of the functional form assumptions of our regression specifications.

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 four different approaches. Our baseline standard errors are based on the standard bootstrap procedure (using 10,000 iterations). Additionally, whenever we include the dummy for the presence of a state e-filing program, we report p-values from the block bootstrap procedure,[42] the p-values based on the robust ("sandwich") covariance matrix estimator that allows for an unknown form of heteroskedasticity, 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) and p-values from the randomization test.[43] 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.

5 Results: Effects on Electronic Filing and EITC Claims

5.1 Electronic Filing

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 1a, 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 show the results of the effect of the introduction of electronic filing programs for state taxes on the number of federal tax returns filed electronically estimated using specification of the type described by equation 2 in Table 2a (using shares) and Table 2b (using logs). In column (1) we show a specification allowing for the set of four post-reform dummy variables, while in column (2) we show the results of a specification that further restricts the effect of

[42] 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.

[43] We implement the randomization test by randomly reassigning the actual dates of treatment across states and re-estimating the same specification. The p-value is the fraction of estimates (based on 10000 draws) with the absolute value that was higher than the baseline one. The randomization p-values address the critique of Bertrand et al. (2004) head on by showing the fraction of similar (artificial) experiments that would result in even higher estimates. While this approach can mechanically assure the right size of the test, its results cannot be interpreted in a standard way. See Kennedy (1995) for the discussion of their use in econometrics.

the introduction of state electronic filing to be constant over time. Columns (3) and (4) are similar to (1) and (2) but allow for state specific linear time trends in electronic filing. In our preferred specifications which include state-specific time trends, we find that there is a large and statistically significant increase in the number of federal electronic filers following the introduction of such state programs.[44]

The share specification indicates that (in an average state) an additional 0.7 to 1 percentage points of the population files electronically, depending on whether or not state-specific time trends are included. The log specification that allows for linear trends predicts a 10% increase in e-filing following the introduction of the program. These effects are large and statistically significant.

In columns 5-8 of tables 2a and 2b, we run the same specifications as in columns 1-4 but we restrict ourselves only to those states which have at least three years of post treatment data. The results from these specifications in which the first three $statefiling$ dummies are identified using a constant number of states, are very similar to our earlier results. They suggest that our findings are not driven by compositional changes that might arise in an unbalanced panel. The results with state-specific time trends are very close to the ones in the baseline specification. Leaving out state trends has only a minor effect on the estimated coefficients. In the final specification we additionally include states without the income tax.[45] Given that the treatment for these states does not vary during the period, these additional observations are solely used to identify the year effects. This extension has no impact on the estimated coefficients that remain significant and very close to the values estimated based on the sample of states with income tax.

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. One technical observation is that state-specific time trends play an important role in the log-specification. This is not particularly surprising because it is certainly possible that the number of e-filers grows at a different speed in different states either because of differences in the rates of population growth or because of differences in the composition of population that affect the rate of adoption of this approach. We pursue our further analysis while allowing for state-specific trends in all of our specifications.[46]

Table 3 presents the results for other income categories. The effect of electronic filing is

[44] In fact of the specifications that we discuss in tables 2a and 2b the only insignificant result is the regression for the logs that do not allow for state-specific time trends. This is the most questionable specification because electronic filing is heavily trending during this period and assuming no differences in trends across states is not appealing.

[45] It is easy to verify that the results must be identical regardless of whether these states are assumed to be always or never treated.

[46] This approach occasionally makes a difference for our results in log specifications that involve dependent variables related to e-filing. The case for using state-specific time trends in the share specifications is weaker and results with and without them are quite similar.

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 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. The informational aspects that we highlighted in the discussion of the tax preparation industry have very limited applicability in the context of higher income individuals who file their taxes anyway. Finally, the alternative methods of e-filing, such as filing through own software from home in the mid-1990s or using the Internet in the late 1990s, are most likely driven by other considerations such as being technology savvy (Goolsbee, 2002).

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 the 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.

5.2 EITC Claims

Next we turn to the impact on EITC claims. Tables 4a and 4b present results for the share and log specifications respectively. These tables follow the same pattern as the tables for electronic filing. The results show the existence of a large positive effect of our treatment on the number of people claiming EITC on their federal returns and, similarly to the effect on electronic filing, this effect persists for the first four years after implementation of these programs. These are the main results of this paper. The following analysis will be devoted to making sure that they are robust and to understanding the economic responses behind them.

The results in Tables 4a are only slightly sensitive to the inclusion of state-specific time trends and trends play no role in the log specification presented in Table 4b.[47] The results using a "balanced" panel of all states are very similar to the baseline ones. The

[47]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

final specification shows the instrumental variable estimate of the effect of e-filing on EITC participation using the four *statefiling* dummies as instruments. The instrumental variable estimate in the share specification has a straightforward interpretation: 77% of an increase in e-filing due to the introduction of e-filing programs corresponds to new EITC participants. The estimate from the log specification suggest that between 1988 and 1999 a 1% increase in the number of e-filers leads, on average, to a 1% increase in the number of EITC claims. This estimate should be interpreted as the average treatment effect over the period. It should also be noted that in all specifications in this paper states are weighted equally.[48]

In Table 5, we present the results by income categories. Consistently with the findings for electronic filing, we find no evidence of response in neither the 0-10K category nor 20-30K category. We also consider the impact of the treatment on the number of individuals not receiving EITC. We find that the number of such individuals (in the log specification) does not respond to the treatment while their share as a fraction of the total population appears to decrease (albeit the coefficients are not significant). This suggests that the increase in EITC participation that we find is driven by new filers rather than by the existing regular filers who were previously unaware of the program.[49]

Since our identification strategy relies on the differential timing of the implementation of these programs, we also added four leads 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, which use our preferred specification that includes state-specific time trends from the unbalanced panel, can be easily presented graphically. Figures 5a and 5b plots the coefficients from the regressions using electronic filing as the dependent variable both for the log and the share specification, while Figure 6a and 6b does the same for EITC claims. 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 percent level) for these coefficients from the regressions. Graphs for electronic filing show estimates for the whole population, the 0-30K, 0-10K and 10-30K category. There is a clear pattern of the effect at time zero for both 10-30K and 0-30K categories, and although confidence intervals are generally large (unsurprisingly, 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 percent 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 Figures 6a and 6b, which plot EITC claims around program implementation. Despite relatively wide confidence intervals, the pattern of an

[48] We found some evidence that the effect is stronger for smaller states.

[49] Because the denominator represents the sum of filers in all income categories, the presence of new filers increases the denominator and, if the number of non-recipients of EITC stays constant, it should decrease their share.

increase in EITC claims following time zero is visible for the 10-30K category and, somewhat less clearly, for the 0-30K category. There is no evidence of a similar response for either the 0-10K category or the non-EITC participants. 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 Tables 6a and 6b we present a few additional specification checks. A few major events affecting the low-income individuals took place in the 1990s. The EITC was expanded twice during the 1990s and in 1996 the welfare system was overhauled. Aggregate changes such as the expansions of the EITC in 1991 and 1994 should be in principle captured by year effects. We test whether it is so. EITC benefits were increased as of 1991. The first column shows our specification estimated using observations starting in 1991. Despite reduced number of observations, the estimated coefficients on e-filing and EITC claims remain significant and of similar magnitude as those estimated using the whole sample.[50] The following two columns present two attempts to control for the impact of the 1996 welfare reform. States designed their own programs meeting TANF requirements between September of 1996 and March of 1997. These changes likely affected labor supply and participation and therefore EITC eligibility of low income individuals (Blank, 2002). These effects were state-specific and therefore may not be appropriately controlled for by year fixed effects. We re-estimate our baseline specification on the pre-1997 observations (column 2) and, as an alternative, we include the interactions of state fixed effects with the post-1996 dummy that are supposed to control for state-specific shifts in the number of eligible EITC claimants (column 3). Neither of these changes affected our estimates. Finally, 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. We have re-estimated our specifications while controlling for the pre-1997 presence of a state waiver (column 4). Our results were virtually unaffected even though the state waiver dummy had an effect on EITC participation but not on e-filing. The introduction of waivers is a relevant consideration for studying EITC participation, but appears orthogonal to the introduction of state e-filing changes.

Two major EITC-related events took place in 1994: the program was expanded and the IRS undertook efforts to reduce cheating. In particular, the IRS began to require that taxpayers include the Social Security number of a qualifying child and reduced its cooperation with tax preparers. There is anecdotal evidence that this change significantly reduced the number of (fraudulent) EITC claims, although no strong effect on EITC participation can be detected in Table 1a. However, as discussed earlier, this change significantly affected the number of e-filers.

[50]This specification also addresses an earlier concern regarding the existence of a number of zeros in the early years of the sample, since all but one occur before 1991. In another specification that we do not report in the table, we dropped all the cells that have a value of zero and this change also did not affect our main findings.

Since both of these events are aggregate, the inclusion of year fixed effects should in principle control for their impact.[51] In addition we pursue two additional specification checks. In column 5 we allow for state-specific pre-1994 dummies to allow for a flexible regime change in 1994. This change has a relatively minor impact on our estimates. In column 6, we allow for the interaction of the treatment dummy with a pre-1994 dummy to allow for the possibility of a change in the effect of treatment. Given that many opportunities for tax evasion were eliminated in 1994, one might expect that the effect of the treatment is stronger pre-1994 if tax evasion was an important part of the response. The interaction is always insignificant and, if anything, it has a negative impact on EITC participation.[52] This specification provides little support for the importance of tax evasion in our results (although it does not eliminate it as an explanation), but it strongly suggests that, despite their magnitude, the 1994 events are appropriately accounted for in our baseline specification.

In column 7 of Tables 6a and 6b we investigate the relevance of another 1994 event. As a part of the IRS's anti-fraud policy, the use of the Direct Deposit Indicator was suspended and IRS cooperation with the tax preparation industry was limited. This policy began to be relaxed as early as in 1995. As explained earlier, this event increased the riskiness of RALs and therefore might have led to a reduction in the attractiveness of electronic filing services. We eliminated 1994 from our sample to test for the importance of this event and it does not appear to have a quantitatively important effect. We interpret it as indicating that year effects appropriately control for this event (i.e., that its impact did not vary in a way that's correlated with our treatment).

In column 8 we run our basic specification by defining treatment solely on whether and when a state introduced the *Federal/State Electronic Filing Program*. As mentioned earlier, we initially defined our treatment date as the year when a state first introduced *any* state electronic filing program. While for most states the first such program was the *Federal/State Electronic Filing Program* implemented in a uniform way in collaboration with the IRS, a number of states (California, Illinois, Maryland, Massachusetts and Minnesota) first introduced their own independent programs separate from the IRS, while Maine's first state electronic income tax filing program was TeleFiling. The results, based on these particular IRS sponsored programs that arguably did not differ across states, are again very similar to the other columns of the tables. In column 9, 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 10 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 and TeleFiling could only be used for non-complicated returns and did not allow for claiming

[51] We attempted splitting the sample in 1994. These results were insignificant with point estimates that were different from our baseline for both subsamples. However, both of them are based on only six years of data and, furthermore, few states were treated post-1994.

[52] The negative sign is consistent with the possibility that the lower number of e-filers pre-1994 corresponds to a weaker effect of the introduction of state e-filing on EITC claims pre-1994.

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), we dropped particular years or states from our sample and found the main results unaffected by these changes. 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 socio-economic conditions in a way correlated with the timing of our treatment.

In Table 7, we present a number of additional specifications intended to verify if the pattern of increases in EITC claims is consistent with the electronic filing story. To do so, we divide EITC recipients using two criteria: whether one used electronic filing and whether a professional prepared the tax return. Any of these four categories is possible. One should be reminded here that *the electronic filing with a self-prepared return* category still corresponds to using a tax professional. In fact, such behavior is very common and one could argue that this type of filing behavior should be most appealing to low income individuals who want to minimize their cost but are interested in a quick refund. Our data does not contain information about using a paid tax preparer for 1989 and 1990, and in 1988 there were very few e-filers; therefore, once the group of e-filers is split by the tax preparation status, we end up with many zeros. Therefore, we pursue this analysis starting with 1991.

The estimates for electronic filing and EITC based on this restricted sample which starts in 1991 are presented in columns 1 and 2 of Table 7 and are very close to those based on the full sample. In the following two columns of this table (3 and 4) we confirm that the whole growth in EITC claims is due to e-filers. This provides further evidence consistent with interpreting our estimates as the causal effect of changes in state electronic filing programs. In columns 5 and 6 of Table 7 we decompose the effects by tax preparation status. There is some evidence that the main source of growth in EITC claims are taxpayers who do not rely on a paid tax preparer to fill out their tax return. As discussed earlier, this category includes taxpayers who are choosing to only file electronically and who are an essential part of our story. The remaining four columns present further decomposition by splitting the sample four-ways. There is statistically significant evidence (in columns 7 of Table 7) of the growth in EITC claims for the professional-help and e-filing group. Not surprisingly, in column 8 of the same table, the number of taxpayers who use a paid tax preparer but choose not to file electronically appears to decline (although the estimates are insignificant

in both specifications), suggesting that the introduction of state filing programs also might have had an effect on how the existing customers of paid tax preparers file their taxes. The estimate of an increase in the number of taxpayers who e-file and self-prepare their tax returns is positive but insignificant. In the final column of Table 7, there is no statistically significant evidence of an increase in EITC claims among taxpayers who file taxes fully on their own.

If one took the (insignificant) coefficient of 0.225 in the share specification seriously, a back-of-the-envelope calculation suggests that the spill-over effect (the informational/network effect on taxpayers who learn about EITC but do not use a paid professional) was approximately $0.225/0.752 \approx 30\%$.[53]

6 Discussion

6.1 The source of new claims

While our analysis so far has focused mainly on identifying the effect on electronic filing and EITC claims of the introduction of state electronic filing, we did not address a number of questions that are crucial for the understanding and interpretation of our results: (1) Are they new filers of taxes or regular filers of taxes who did not claim EITC before? (2) More generally, who are the new EITC claimers? (3) Are the new filers from families who qualify for the program but for some reason did not claim the benefits they were entitled to, or are they simply part of the potentially large number of cheaters who did not comply with the eligibility criteria of the program?

We have seen before that there is no evidence of a reduction in the number of EITC recipients in response to state electronic filing programs. The direct way of testing whether the new EITC claimers are also new to filing federal taxes is to run our regression analysis using the number of people filing. Simultaneously, this approach will allow us to address the possibility of a behavioral response resulting in shifting across categories. Table 8 contains the results of this exercise. The results are somewhat imprecise but supportive of the hypothesis that the increase corresponds to new filers. The number of total tax recipients in the 10-30K category increases by 3%, and is statistically significant at 10% level. This is consistent with the hypothesis that the whole growth in claims is due to increased filing by non-filers: given that the number of EITC recipients in this category varied between twenty-one and twenty-six percent during the period and given the baseline estimate of the impact on EITC participation of 12.6%, the corresponding increase in the number of tax returns if all EITC claims are external should be between 2.6% and 3.3%.

There is also no evidence of a decrease in the number of filers in the 0-10K category

[53] We also estimated the dynamics of the responses by category using specifications with leads and lags. Estimates for these specifications were imprecise but not showing any evidence of pre-trends that would suggest a misspecification.

refuting the possibility that the whole growth in EITC claims was just driven by the response of lower income individuals motivated by the possibility of obtaining higher EITC benefits and shifting to the 10-30K category. As a consistency check, there is also no evidence that the number of returns for taxpayers with incomes above the EITC phase-out limit responded to our treatment. The estimates based on the share specification are insignificant but consistent with the ones based on logs. These results do not exclude the possibility that there was indeed some behavioral response stimulated by the increased awareness of EITC due to the introduction of electronic filing opportunities. The 10-20K category contains the region where benefits are maximized. As a result, it is possible that claims increased for non-filers who would belong to either of the three income categories, but that these non-filers simultaneously responded by adjusting their incomes so that they fall into the maximum benefit range.

The difference between the share estimates for the 10-30K category (approximately 0.6) in Table 3 and the total effect on EITC claims (approximately 0.9), measures the number of new EITC recipients who used to file their tax returns: the difference is about 0.3 and it is 1/3 of the total increase in EITC claims. Given the lack of precision of estimates in Table 3 this is only suggestive of the size of this type of response.

6.2 Cheating

Any increase in participation may be reflective of increased cheating. There are two different margins on which cheating could operate. Cheaters may have come from either the pool of filers or the pool of non-filers. Certain types of cheating by usual filers have observable implications. Claiming ineligible children is the major source of noncompliance (Holtzblatt and McCubbin, 2003). It may increase the number of children reported on the tax returns. Another important aspect of non-compliance involves changing filing status. For example, filing separately from the spouse as single or head of household (married filing separately are not eligible for EITC) in many cases would lead to an increase in the EITC benefits, but this is illegal. Nevertheless, to the extent that it takes place, one would expect, ceteris paribus, a decrease in the number of joint filers.

We have already attempted two specification checks related to cheating in Tables 6a and 6b: we found no evidence of a differential effect surrounding or following the 1994 change in enforcement policies. However, this is not a definitive finding, because it is possible that these changes had a relatively uniform effect on tax fraud regardless of the treatment status of the states.

We attempt to shed further light on the importance of cheating. In Table 9 we concentrate on the possibility of adjusting marital status in order to claim higher benefits. In the first column we compare estimates for the total number of returns with the estimates for the total number of *people* constructed by double-counting joint tax returns. If there is cheating by manipulation of the filing status, we would expect the number of returns to in-

crease by more than the number of people. In the extreme case the number of people could stay constant with the number of returns increasing. We find no such evidence: estimates for the number of returns and the number of bodies are indistinguishable. In the following three columns we decompose the change in the number of returns into the response of different filing categories to see if there are possible compositional shifts. The only category that appears to be strongly responding are heads of households (although estimates are not statistically significant), while none of the other coefficients is negative. Therefore, there is little support for the possibility that taxpayers adjusted their filing status to claim higher benefits. The following three columns show the results for the EITC recipients only. There is evidence there that there was a significant growth in both the number of joint filers and heads of household. There is no similar evidence for single EITC recipients, but it should be noted that this group is somewhat peculiar: single (non head-of-household) filers can receive EITC while in the 10-30K income category only if they have qualifying children. Therefore, these are taxpayers who claim qualifying children for EITC purposes but not a dependent exemption. One might have expected that cheaters would pursue this path, but there is no evidence that this actually happened.

The comparison of estimates for the number of returns and the number of recipients by marital status categories is revealing. While the number of EITC recipients who are jointly filing increases, the total number of joint filers does not. Given no evidence of endogenous changes in marital status, it suggests therefore that the new jointly filing EITC recipients used to file in the past. This is in contrast to the estimates for heads of households: there is a similar increase in the total number of taxpayers claiming this status and in the number of EITC recipients that claim it. As a result, it appears that the EITC response may in fact reflect both increased filing and learning by existing filers: the first effect applies to singles with children while the latter is the domain of joint filers.

In Table 10 we investigate the possibility of cheating by claiming non-qualifying children. The ideal measure for this purpose would be the number of qualifying children for EITC purposes, but this measure is not available in our data. As mentioned earlier, different qualifying criteria are used for EITC and tax exemption purposes. Nevertheless, if there is systematic and large scale over-reporting of ineligible children we might expect this to be accompanied by overstating dependent exemptions as well.

In the first column of Table 10, we investigate the effect of the treatment on the total number of child exemptions. It is somewhat imprecisely estimated but relatively large (and borderline significant in the log specification). The relevant reference point for this coefficient is the increase in the number of tax returns reported in Table 9. The comparison of the points estimates in the share specification suggests that the number of children increases by slightly more than the number of tax returns. In other words, the additional filers appear to have on average slightly more than one child. The log specifications suggest that the number of children increases at the rate twice as large as the number of returns.

The following two columns show the impact on the number of children by the EITC

status. These results might be interpreted as suggesting that the increased EITC participation involved some reshuffling of children, potentially corresponding to cheating. The estimated increase in the total number of children of EITC recipients is of the order of 1% of the total number of returns. This estimate is higher than the increase in the total number of child exemptions and, correspondingly, we observe a negative coefficient on the number of children among the non-EITC recipients. Recall that the total number of non-EITC recipients decreased slightly in the share specification and was flat in the log specification (Table 5). A decrease in the number of children of non-EITC recipients would have to then correspond to shifting of children to EITC recipients. Even though this is suggestive of possible cheating, all of these estimates are insignificant and small relative to the increase in the number of children among EITC recipients.

Our evidence on cheating can be summarized as follows. We did not find any evidence of cheating by adjusting the marital status. However, the decomposition by marital status allowed us to conclude that heads of households are the main pool of nonfilers contributing to increased EITC participation. We also find some evidence of a smaller but statistically significant effect on participation of joint filers who used to file their tax returns. We also investigated whether there is any evidence of cheating by manipulating the number of reported children. There is very weak evidence that the number of children of non-EITC recipients decreases somewhat even though the number of non-EITC recipients is stable (in the log specification). This is suggestive of potential shifting of children from non-EITC recipients to EITC recipients. However these results are not statistically significant. Stressing this caveat, the point estimates would imply that 1/3 of newly reported tax exemptions by EITC recipients are for children who would have been claimed by someone on a tax return anyway. Recalling, however, that we have already seen that there is a weak evidence of the existence of a group of joint filers who used to file and yet did not apply for the EITC before state electronic filing programs were introduced, the reduction in the number of children by non EITC recipients may be simply due to this group of filers changing EITC status.

In the last three columns of Table 10, we split the sample of EITC recipients by the number of children. We find that the new claimants are mostly taxpayers with only one child. We conclude therefore by observing that a typical new EITC participant is a single taxpayer with one child filing as a head of household who used not to file before, arguably the type of person that government programs are trying to reach. While these findings imply that, based on the available observable characteristics, the new filers of EITC claims are fairly similar to existing filers, we cannot entirely rule out the possibility that a portion of these new filers are in fact cheaters who did not comply with the criteria of the federal EITC program.

7 Conclusions and Implications

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. Furthermore, we suggest that this effect have been driven by a response of the tax preparation industry. In our analysis we provided a large 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 1990's.

Additionally, we find that increased EITC participation draws mostly from the pool of non-filers: individuals who otherwise would not file their tax returns. We do not find any major evidence suggesting that cheating is an important part of the effects that we estimate. While fraud certainly still exists in the system, it does not appear that cheating was further stimulated by *state* electronic filing opportunities.

We make three additional observations that follow from our results. First, while our results point to a potential benefit of the decentralized tax preparation industry, the use of RAL's and the role of tax preparation industry remain controversial, because they reduce the actual benefits received by low-income individuals. We do not provide any evidence related to the costs of their activity, but only point out that these costs should be evaluated against benefits from increased participation and the costs of alternative means of providing information about EITC, such as government-sponsored outreach efforts.[54] A full evaluation of the role of the tax preparation industry would require additional information, such as medium and long term responses within the tax preparation industry, the change of eligibility criteria over time as well as the size and composition of uninformed taxpayers.

Second, the results illustrate that complexity is a serious issue in the program and tax design. More precise targeting of benefits requires increased complexity, but increased complexity will make it less likely that intended recipients would be able to benefit. In this sense, concerns about complexity (separately from traditional efficiency arguments) may limit the extent of effective redistribution. On the positive side, our results suggest that such trade-offs are likely sensitive to institutional and technological environments that are subject to change.

Finally, our analysis provides a cautionary note to the studies that rely on repeated cross-sections of taxpayers. We found that it is not prudent to assume that taxpayers in every year are drawn from the same distribution. In particular, our results suggest that during the 1990s the lower end of the distribution of individuals who are captured by tax statistics expanded.

[54] See Aizer (2004) for the evidence of effectiveness of state-run outreach and advertising campaigns in the context of Medicaid.

References

Aizer, Anna, "Advertising, Medicaid and Child Health," March 2004. Brown University, mimeo.

Akerlof, George A., "The Economics of "Tagging" as Applied to Optimal Income Tax, Welfare Programs, and Manpower Planning," *American Economic Review*, March 1978, *68* (1), 8–19.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, "How Much Should We Trust Differences-in-Differences Estimates?," *Quarterly Journal of Economics*, February 2004, *119* (1), 249–275.

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

Blank, Rebecca M., "Evaluating Welfare Reform in the United States," *Journal of Economic Literature*, December 2002, *40* (4), 1105–1166.

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

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

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

Erard, Brian and Chih Chin Ho, "Searching for Ghosts: Who Are the Nonfilers and How Much Tax Do They Owe?," *Journal of Public Economics*, July 2001, *81* (1), 25–50.

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

Goolsbee, Austan, "The Turbo Tax Revolution? Evaluating the Ability of Technology to Solve the Tax Complexity Dilemma," January 2002. University of Chicago, mimeo.

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

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

Horowitz, Joel L., "The Bootstrap," in James J. Heckman and Edward Leamer, eds., *Handbook of Econometrics*, Vol. 5, Amsterdam; New York: Elsevier/North Holland, 2000.

Hotz, V. Joseph and John Karl Scholz, "The Earned Income Tax Credit," in Robert A. Moffitt, ed., *Means-Tested Transfer Programs in the United States*, University of Chicago Press, 2003.

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

Kennedy, Peter E., "Randomization Tests in Econometrics," *Journal of Business and Economic Statistics*, January 1995, *13* (1), 85–94.

Liebman, Jeffrey B., "The EITC compliance problem," *Poverty Research News*, Summer 1998, *2* (3). `http://www.jcpr.org/98summer/article3.html`.

_ , "Who are the Ineligible EITC Recipients?," *National Tax Journal*, December 2000, *53* (4), 1165–1185. Part 2.

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

McCubbin, Janet, "EITC Noncompliance: The Determinants of the Misreporting of Children," *National Tax Journal*, December 2000, *53* (4), 1135–1164. Part 2.

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

Scholz, John Karl, "The Earned Income Tax Credit: Participation, Compliance and Anti-poverty Effectiveness," *National Tax Journal*, March 1994, *48* (1), 59–81.

Wu, Chi Chi and Jean Ann Fox, "The High Cost of Quick Tax Money: Tax Preparation, 'Instant Refund' Loans, and Check Cashing Fees Target the Working Poor," Report, National Consumer Law Center January 2003. Online at `http://www.consumerfed.org/2003_RAL_report.pdf`.

Table 1a: EITC, E-Filers and Professionally Prepared Returns by Income Group and Year

	% EITC				% E-filers				% Professionally Prepared Return			
	Total	0-10K	10-30K	>30K	Total	0-10K	10-30K	>30K	Total	0-10K	10-30K	>30K
1988	0.10	0.13	0.21	0.00	0.01	0.01	0.02	0.01	0.47	0.34	0.45	0.57
1989	0.11	0.14	0.21	0.00	0.04	0.03	0.06	0.03				
1990	0.11	0.15	0.22	0.00	0.08	0.07	0.11	0.06				
1991	0.12	0.17	0.23	0.00	0.10	0.11	0.14	0.06	0.49	0.40	0.48	0.55
1992	0.13	0.17	0.24	0.00	0.11	0.11	0.16	0.07	0.50	0.41	0.49	0.55
1993	0.13	0.19	0.25	0.00	0.13	0.13	0.18	0.08	0.50	0.41	0.49	0.55
1994	0.17	0.33	0.25	0.00	0.10	0.10	0.14	0.07	0.50	0.41	0.50	0.55
1995	0.17	0.32	0.26	0.00	0.13	0.15	0.16	0.08	0.50	0.43	0.49	0.56
1996	0.17	0.32	0.26	0.00	0.16	0.19	0.21	0.10	0.51	0.42	0.51	0.56
1997	0.16	0.31	0.26	0.00	0.20	0.23	0.26	0.13	0.52	0.44	0.52	0.57
1998	0.16	0.32	0.26	0.00	0.23	0.25	0.30	0.17	0.53	0.45	0.53	0.57
1999	0.15	0.30	0.26	0.00	0.28	0.28	0.35	0.23	0.54	0.47	0.55	0.58

Table 1b: E-filing and Professionally Prepared Returns among EITC recipients

	% E-filers			% Professionally Prepared Returns			% E-filers and Professional Returns		
	Total	0-10K	10-30K	Total	0-10K	10-30K	Total	0-10K	10-30K
1988	0.04	0.04	0.04	0.43	0.39	0.45	0.03	0.03	0.02
1989	0.13	0.11	0.13						
1990	0.25	0.24	0.24						
1991	0.36	0.36	0.33	0.51	0.53	0.51	0.23	0.24	0.21
1992	0.40	0.39	0.38	0.55	0.52	0.55	0.27	0.26	0.25
1993	0.42	0.39	0.40	0.55	0.53	0.56	0.29	0.26	0.28
1994	0.26	0.19	0.30	0.52	0.46	0.56	0.19	0.14	0.22
1995	0.29	0.25	0.30	0.54	0.52	0.55	0.21	0.19	0.22
1996	0.36	0.33	0.37	0.57	0.53	0.60	0.28	0.24	0.30
1997	0.42	0.38	0.42	0.60	0.57	0.62	0.34	0.30	0.35
1998	0.45	0.42	0.47	0.62	0.57	0.65	0.37	0.33	0.39
1999	0.54	0.49	0.56	0.67	0.63	0.69	0.45	0.40	0.47

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

Table 1c: Summary Statistics, by State (1988-1999)

	Year of			% of returns in		% of returns with		
	State EFI	Telefile	StateFed	0-10K	10-30K	EITC	E-file	Prof. help
AL	1997	none	1997	25.70	38.09	23.56	18.71	52.58
AK	no tax	no tax	no tax	40.46	23.38	8.06	8.93	40.03
AZ	1997	none	1997	22.15	37.85	16.45	12.69	53.32
AR	1994	1997	1994	24.92	41.34	21.57	16.72	57.33
CA	1994	1996	never	21.33	34.02	15.22	7.88	57.00
CO	1993	none	1993	20.32	32.96	11.15	11.05	50.09
CT	1993	1998	1993	19.88	30.98	7.69	9.95	46.04
DE	1993	1999	1993	18.33	33.21	11.39	12.04	42.40
DC	1996	2000	1996	14.75	42.45	14.13	12.20	42.21
FL	no tax	no tax	no tax	22.73	38.86	15.52	15.80	50.03
GA	1994	2001	1994	22.24	36.17	19.05	18.58	51.41
HI	2001	none	2001	19.78	36.82	8.07	7.18	53.31
ID	1993	1999	1993	26.37	40.44	17.31	12.11	48.05
IL	1990	1993	1995	23.28	32.28	12.18	14.44	51.25
IN	1992	1999	1992	22.75	33.51	11.76	18.33	50.44
IA	1993	none	1993	23.90	35.28	9.00	13.30	62.74
KS	1991	1996	1991	23.31	33.63	12.18	14.15	57.94
KY	1992	1999	1992	24.60	36.79	15.53	17.75	55.68
LA	1992	1996	1992	25.38	38.04	24.75	17.90	49.72
ME	2000	1996	never	24.25	37.07	10.03	10.82	39.04
MD	1990	2001	1995	19.64	32.59	11.23	11.19	44.98
MA	1992	1995	never	21.05	32.00	7.41	9.58	47.48
MI	1992	1999	1992	23.88	31.56	10.87	12.51	50.03
MN	1989	1998	never	23.69	32.56	7.87	10.70	50.85
MS	1992	1995	1992	25.11	43.19	30.53	20.27	47.46
MO	1993	1997	1993	24.80	35.70	14.59	15.17	56.55
MT	1994	1997	1994	26.53	39.35	13.59	13.08	57.38
NE	1993	1998	1993	23.03	35.44	12.05	10.67	57.30
NV	no tax	no tax	no tax	18.28	39.43	13.80	13.56	50.24
NH	no tax	no tax	no tax	21.56	31.40	7.71	13.00	40.62
NJ	1994	1996	1994	19.72	29.98	10.20	9.48	53.77
NM	1992	none	1992	28.54	36.45	24.44	16.88	49.66
NY	1992	none	1992	21.22	33.76	12.59	8.76	55.78
NC	1991	none	1991	23.27	38.21	17.08	17.48	51.89
ND	1999	none	1999	23.91	36.01	9.15	10.71	60.88
OH	1999	1997	1999	23.11	36.02	10.72	14.01	48.11
OK	1992	2001	1992	24.52	38.56	16.97	14.30	57.54
OR	1993	2001	1993	20.15	35.27	10.16	8.22	47.71
PA	1995	1998	1995	23.29	35.36	10.56	10.28	45.35
RI	1994	none	1994	20.35	34.59	11.31	10.35	50.88
SC	1990	1997	1990	24.99	40.63	19.78	23.78	53.33
SD	no tax	no tax	no tax	25.74	40.26	11.92	14.99	50.58
TN	no tax	no tax	no tax	24.65	37.48	18.57	20.01	49.23
TX	no tax	no tax	no tax	24.93	36.70	19.94	16.67	45.33
UT	1992	1998	1992	21.94	34.52	11.22	11.90	45.62
VT	2001	none	2001	24.95	32.37	10.79	6.72	43.37
VA	1994	2000	1994	19.90	33.35	11.27	13.29	43.85
WA	no tax	no tax	no tax	19.36	32.52	9.24	11.50	42.33
WV	1991	2001	1991	23.92	40.57	17.04	14.57	42.33
WI	1991	1998	1991	23.37	34.56	9.71	12.63	48.97
WY	no tax	no tax	no tax	25.80	35.03	9.59	16.07	50.89
Total				23.17	35.66	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).

Table 2a: Regression analysis: Electronic Filing (shares, 10-30K)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
State EFI		0.716		1.049		1.054		1.175	0.983
		(0.378)*		(0.421)**		(0.421)**		(0.490)**	(0.416)**
State EFI (1)	0.808		0.971		1.011		1.119		
	(0.301)***		(0.348)***		(0.351)***		(0.412)***		
State EFI (2)	0.709		0.908		0.924		1.046		
	(0.478)		(0.576)		(0.562)		(0.662)		
State EFI (3)	0.721		0.943		1.046		1.176		
	(0.516)		(0.682)		(0.646)		(0.752)		
State EFI (4)	0.202		0.408		0.631		0.706		
	(0.592)		(0.786)		(0.785)		(0.870)		
Robust		0.008		0.000		0.001		0.001	0.001
Clustered		0.088		0.034		0.028		0.043	0.039
Bootstrap		0.058		0.015		0.015		0.020	0.020
Randomiz.		0.040		0.006		0.005		0.003	0.007
N	504	504	504	504	420	420	420	420	612
States	42	42	42	42	35	35	35	35	51
Year dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes
State dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes
State trends	no	no	yes	yes	no	no	yes	yes	yes

Notes: The analysis is for the period 1988-1999. The dependent variables are measured in shares and refer to the 10-30K income category. All regressions include year and state fixed effects and some specifications also include state specific time trends. State EFI refers to the average treatment effect, while State EFI (1) - State EFI (4) indicate the effect in the first four years after program implementation. 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 robust standard errors, clustered standard errors (at the state level) as well as the randomization test. $***$, $**$ and $*$ indicate statistical significance at the 1, 5 and 10 percent levels, respectively.

Table 2b: Regression analysis: Electronic Filing (logs, 10-30K)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
State EFI		0.035		0.108		0.132		0.138	0.090
		(0.070)		(0.057)*		(0.074)*		(0.066)**	(0.056)
State EFI (1)	0.054		0.108		0.134		0.141		
	(0.057)		(0.050)**		(0.065)**		(0.058)**		
State EFI (2)	0.006		0.080		0.107		0.116		
	(0.079)		(0.076)		(0.094)		(0.089)		
State EFI (3)	0.066		0.152		0.192		0.204		
	(0.089)		(0.090)*		(0.105)*		(0.101)**		
State EFI (4)	-0.026		0.095		0.146		0.163		
	(0.105)		(0.104)		(0.110)		(0.114)		
Robust		0.443		0.016		0.008		0.007	0.042
Clustered		0.645		0.106		0.113		0.078	0.155
Bootstrap		0.619		0.062		0.076		0.040	0.109
Randomiz.		0.518		0.056		0.034		0.026	0.105
N	504	504	504	504	420	420	420	420	612
States	42	42	42	42	35	35	35	35	51
Year dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes
State dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes
State trends	no	no	yes	yes	no	no	yes	yes	yes

Notes: The analysis is for the period 1988-1999. The dependent variables are measured in logs and refer to the 10-30K income category. All regressions include year and state fixed effects and some specifications also include state specific time trends. State EFI refers to the average treatment effect, while State EFI (1) State EFI (4) indicate the effect in the first four years after program implementation. 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 robust standard errors, clustered standard errors (at the state level) as well as the randomization test. $***$, $**$ and $*$ indicate statistical significance at the 1, 5 and 10 percent levels, respectively.

Table 3: Regressions: Impact on Electronic Filing by Income Group

	Shares						
	All	0-30K	0-10K	10-30K	10-20K	20-30K	30-100K
State EFI	0.632	0.949	-0.100	1.049	0.802	0.247	-0.191
	(0.520)	(0.518)*	(0.234)	(0.421)**	(0.331)**	(0.172)	(0.196)
Robust	0.149	0.013	0.600	0.000	0.001	0.131	0.292
Clustered	0.289	0.113	0.708	0.034	0.037	0.215	0.380
Bootstrap	0.222	0.064	0.669	0.015	0.016	0.155	0.334
Randomiz.	0.232	0.053	0.657	0.006	0.004	0.216	0.394
	Logs						
	All	0-30K	0-10K	10-30K	10-20K	20-30K	30-100K
State EFI	0.052	0.080	0.014	0.108	0.192	0.027	0.013
	(0.040)	(0.051)	(0.066)	(0.057)*	(0.091)**	(0.053)	(0.054)
Robust	0.175	0.062	0.799	0.016	0.002	0.574	0.784
Clustered	0.250	0.177	0.847	0.106	0.068	0.654	0.827
Bootstrap	0.187	0.119	0.826	0.062	0.034	0.605	0.811
Randomiz.	0.288	0.140	0.829	0.056	0.005	0.662	0.827
N	504	504	504	504	504	504	504
States	42	42	42	42	42	42	42
Year dummies	yes	yes	yes	yes	yes	yes	yes
State dummies	yes	yes	yes	yes	yes	yes	yes
State trends	yes	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. 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 robust standard errors, clustered standard errors (at the state level) as well as the randomization test. $***$, $**$ and $*$ indicate statistical significance at the 1, 5 and 10 percent levels, respectively.

Table 4a: Regression analysis: Participation in Earned Income Tax Credit (shares, 10-30K)

										IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	Second
State EFI		0.760		0.901		1.002		1.135	0.903	
		(0.265)***		(0.367)**		(0.365)***		(0.407)***	(0.362)**	
Electronic filers										0.768
										(0.042)***
State EFI (1)	0.660		0.739		0.800		0.954			
	(0.259)***		(0.328)**		(0.319)**		(0.350)***			
State EFI (2)	0.855		1.004		0.993		1.242			
	(0.349)**		(0.467)**		(0.474)**		(0.537)**			
State EFI (3)	0.999		1.287		1.138		1.486			
	(0.448)**		(0.594)**		(0.573)**		(0.661)**			
State EFI (4)	0.401		0.982		0.424		1.054			
	(0.373)		(0.551)*		(0.555)		(0.618)*			
Robust		0.003		0.004		0.003		0.002	0.003	
Clustered		0.011		0.035		0.016		0.019	0.029	
Bootstrap		0.004		0.015		0.008		0.006	0.012	
Randomiz.		0.014		0.007		0.006		0.002	0.005	
N	504	504	504	504	420	420	420	420	612	504
States	42	42	42	42	35	35	35	35	51	42
Year dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
State dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
State trends	no	no	yes	yes	no	no	yes	yes	yes	yes

Notes: The analysis is for the period 1988-1999. The dependent variables are measured in shares and refer to the 10-30K income category. All regressions include year and state fixed effects and some specifications also include state specific time trends. State EFI refers to the average treatment effect, while State EFI (1) - State EFI (4) indicate the effect in the first four years after program implementation. 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 robust standard errors, clustered standard errors (at the state level) as well as the randomization test. $***$, $**$ and $*$ indicate statistical significance at the 1, 5 and 10 percent levels, respectively.

Table 4b: Regression analysis: Participation in Earned Income Tax Credit (logs, 10-30K)

										IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	Second
State EFI		0.121		0.126		0.140		0.154	0.127	
		(0.040)***		(0.050)**		(0.056)**		(0.055)***	(0.048)***	
Electronic filers										1.022
										(0.043)***
State EFI (1)	0.098		0.106		0.119		0.133			
	(0.036)***		(0.043)**		(0.047)**		(0.047)***			
State EFI (2)	0.130		0.145		0.150		0.169			
	(0.049)***		(0.060)**		(0.071)**		(0.069)**			
State EFI (3)	0.164		0.197		0.195		0.219			
	(0.059)***		(0.072)***		(0.083)**		(0.081)***			
State EFI (4)	0.120		0.171		0.146		0.176			
	(0.058)**		(0.066)***		(0.080)*		(0.073)**			
Robust		0.001		0.003		0.002		0.002	0.002	
Clustered		0.007		0.027		0.027		0.019	0.022	
Bootstrap		0.003		0.013		0.011		0.005	0.008	
Randomiz.		0.003		0.004		0.002		0.001	0.003	
N	504	504	504	504	420	420	420	420	612	504
States	42	42	42	42	35	35	35	35	51	42
Year dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
State dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
State trends	no	no	yes	yes	no	no	yes	yes	yes	yes

Notes: The analysis is for the period 1988-1999. The dependent variables are measured in logs and refer to the 10-30K income category. All regressions include year and state fixed effects and some specifications also include state specific time trends. State EFI refers to the average treatment effect, while State EFI (1) State EFI (4) indicate the effect in the first four years after program implementation. 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 robust standard errors, clustered standard errors (at the state level) as well as the randomization test. $***$, $**$ and $*$ indicate statistical significance at the 1, 5 and 10 percent levels, respectively.

Table 5: Regressions: Impact on EITC by Income Group

	Shares						
	0-10K	10-30K	10-20K	20-30K	0-30K	No EITC, 0-30K	No EITC, 10-30K
State EFI	-0.200	0.901	0.642	0.259	0.701	-0.468	-0.285
	(0.220)	(0.367)**	(0.278)**	(0.148)*	(0.456)	(0.482)	(0.476)
Robust	0.414	0.004	0.015	0.149	0.054	0.353	0.561
Clustered	0.422	0.035	0.048	0.124	0.182	0.385	0.590
Bootstrap	0.362	0.015	0.022	0.082	0.125	0.328	0.550
Randomiz.	0.452	0.007	0.015	0.171	0.088	0.371	0.602
	Logs						
State EFI	-0.013	0.126	0.214	0.067	0.067	0.004	0.003
	(0.039)	(0.050)**	(0.073)***	(0.048)	(0.033)**	(0.013)	(0.018)
Robust	0.764	0.003	0.006	0.189	0.019	0.751	0.854
Clustered	0.765	0.027	0.014	0.210	0.080	0.763	0.867
Bootstrap	0.732	0.013	0.005	0.155	0.045	0.735	0.851
Randomiz.	0.798	0.004	0.004	0.254	0.046	0.778	0.879
N	504	504	504	504	504	504	504
States	42	42	42	42	42	42	42
Year dummies	yes	yes	yes	yes	yes	yes	yes
State dummies	yes	yes	yes	yes	yes	yes	yes
State trends	yes	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. 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 robust standard errors, clustered standard errors (at the state level) as well as the randomization test. $***$, $**$ and $*$ indicate statistical significance at the 1, 5 and 10 percent levels, respectively.

Table 6a: Specification checks (shares)

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Electronic Filing	State EFI	1.135	0.972	0.873	1.049	0.898	0.862	1.185		1.050	0.938
		(0.397)***	(0.438)**	(0.416)**	(0.421)**	(0.473)*	(0.542)	(0.460)**		(0.422)**	(0.397)**
	State Waiver				0.052						
					(0.333)						
	Pre-1994 EFI						0.332				
							(0.633)				
	StateFed EFI								1.107		
									(0.441)**		
	State EITC									0.083	
										(0.809)	
	State Telefiling										1.112
											(0.132)***
	Robust	0.001	0.005	0.007	0.000	0.007	0.033	0.000	0.000	0.000	0.001
	Clustered	0.019	0.067	0.088	0.034	0.118	0.160	0.030	0.033	0.035	0.044
	Bootstrap	0.006	0.029	0.038	0.015	0.060	0.111	0.011	0.015	0.015	0.020
	Randomiz.	0.003	0.023	0.033	0.006	0.028	0.088	0.005	0.004	0.006	0.014
Earned Income Tax Credit	State EFI	0.752	0.954	0.849	0.903	0.659	1.064	0.880		0.898	0.895
		(0.417)*	(0.359)***	(0.336)**	(0.358)**	(0.444)	(0.508)**	(0.362)**		(0.366)**	(0.363)**
	State Waiver				0.342						
					(0.269)						
	Pre-1994 EFI						-0.290				
							(0.416)				
	StateFed EFI								0.928		
									(0.386)**		
	State EITC									-0.234	
										(0.391)	
	State Telefiling										0.061
											(0.094)
	Robust	0.046	0.005	0.008	0.004	0.061	0.032	0.007	0.006	0.004	0.004
	Clustered	0.122	0.026	0.037	0.031	0.216	0.066	0.041	0.039	0.035	0.034
	Bootstrap	0.074	0.007	0.013	0.013	0.141	0.035	0.014	0.018	0.015	0.015
	Randomiz.	0.045	0.008	0.013	0.007	0.090	0.016	0.021	0.009	0.007	0.008
	N	378	378	504	504	504	504	462	504	504	504
	States	42	42	42	42	42	42	42	42	42	42

Notes: Notes to previous tables apply here as well. Specifications are as follows: (1) post 1990 data; (2) pre-1997 data; (3) post-1996 state-specific dummies; (4) pre-1996 state welfare waiver present. (5) pre-1994 state-specific dummies; (6) pre-1994 e-filing treatment effect; (7) 1994 excluded; (8) treatment based on timing of State/Fed program; (9) state EITC dummy; (10) Telefiling dummy.

Table 6b: Specification checks (logs)

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Electronic Filing	State EFI	0.123	0.091	0.082	0.108	0.100	0.089	0.125		0.109	0.096
		(0.058)**	(0.062)	(0.058)	(0.058)*	(0.061)*	(0.078)	(0.061)**		(0.059)*	(0.054)*
	State Waiver				-0.015						
					(0.056)						
	Pre-1994 EFI						0.033				
							(0.087)				
	StateFed EFI								0.099		
									(0.057)*		
	State EITC									0.114	
										(0.138)	
	State Telefiling										0.119
											(0.015)***
	Robust	0.011	0.085	0.092	0.016	0.042	0.165	0.008	0.031	0.015	0.028
	Clustered	0.075	0.220	0.240	0.109	0.172	0.305	0.080	0.134	0.113	0.129
	Bootstrap	0.034	0.149	0.158	0.063	0.099	0.250	0.040	0.086	0.065	0.077
	Randomiz.	0.026	0.144	0.153	0.057	0.125	0.208	0.036	0.076	0.053	0.090
Earned Income Tax Credit	State EFI	0.098	0.137	0.121	0.127	0.089	0.160	0.126		0.126	0.126
		(0.061)	(0.054)**	(0.050)**	(0.047)**	(0.060)	(0.077)**	(0.043)***		(0.050)**	(0.049)**
	State Waiver				0.083						
					(0.039)**						
	Pre-1994 EFI						-0.061				
							(0.065)				
	StateFed EFI								0.121		
									(0.052)**		
	State EITC									-0.014	
										(0.050)	
	State Telefiling										0.001
											(0.012)
	Robust	0.056	0.003	0.006	0.002	0.059	0.030	0.002	0.006	0.003	0.003
	Clustered	0.167	0.031	0.043	0.021	0.208	0.063	0.015	0.046	0.028	0.026
	Bootstrap	0.111	0.012	0.016	0.010	0.138	0.037	0.004	0.023	0.013	0.013
	Randomiz.	0.060	0.005	0.008	0.004	0.088	0.008	0.008	0.010	0.004	0.004
	N	378	378	504	504	504	504	462	504	504	504
	States	42	42	42	42	42	42	42	42	42	42

Notes: Notes to previous tables apply here as well. Specifications are as follows: (1) post 1990 data; (2) pre-1997 data; (3) post-1996 state-specific dummies; (4) pre-1996 state welfare waiver present. (5) pre-1994 state-specific dummies; (6) pre-1994 e-filing treatment effect; (7) 1994 excluded; (8) treatment based on timing of State/Fed program; (9) state EITC dummy; (10) Telefiling dummy.

Table 7: EITC Participation: Decomposition of Response

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
			Electronic Filing		Return Prepared by a Professional		Return Prepared by a Professional		Self-Prepared Return	
	EFI	EITC	Yes	No	Yes	No	EFI	No EFI	EFI	No EFI
	Shares, 10-30K									
State EFI	1.135	0.752	0.815	-0.063	0.254	0.498	0.542	-0.287	0.273	0.225
	(0.397)***	(0.417)*	(0.309)***	(0.274)	(0.240)	(0.270)*	(0.241)**	(0.246)	(0.177)	(0.225)
Robust	0.001	0.046	0.003	0.816	0.306	0.028	0.023	0.187	0.064	0.237
Clustered	0.019	0.122	0.029	0.844	0.361	0.117	0.063	0.321	0.191	0.396
Bootstrap	0.006	0.074	0.009	0.820	0.289	0.065	0.023	0.245	0.130	0.323
Randomiz.	0.003	0.045	0.003	0.823	0.341	0.034	0.008	0.150	0.046	0.280
	Logs, 10-30K									
State EFI	0.123	0.098	0.164	-0.000	0.046	0.083	0.156	-0.067	0.113	0.058
	(0.058)**	(0.061)	(0.081)**	(0.045)	(0.049)	(0.068)	(0.078)**	(0.067)	(0.075)	(0.069)
Robust	0.011	0.056	0.022	0.997	0.331	0.121	0.059	0.251	0.090	0.279
Clustered	0.075	0.167	0.083	0.998	0.414	0.292	0.097	0.391	0.193	0.472
Bootstrap	0.034	0.111	0.043	0.997	0.340	0.222	0.050	0.321	0.135	0.410
Randomiz.	0.026	0.060	0.026	0.998	0.385	0.170	0.024	0.249	0.075	0.343
N	378	378	378	378	378	378	378	378	378	378
States	42	42	42	42	42	42	42	42	42	42
Year dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
State dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
State trends	yes	yes	yes	yes	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. 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 robust standard errors, clustered standard errors (at the state level) as well as the randomization test. $***$, $**$ and $*$ indicate statistical significance at the 1, 5 and 10 percent levels, respectively.

Table 8: Number of filers

	Not 0-30K	0-30K	0-10K	10-30K	10-20K	20-30K
			Shares			
State EFI	-0.233	0.233	-0.383	0.616	0.619	-0.003
	(0.491)	(0.491)	(0.404)	(0.468)	(0.390)	(0.427)
Robust	0.572	0.572	0.325	0.157	0.102	0.995
Clustered	0.676	0.676	0.405	0.243	0.166	0.996
Bootstrap	0.645	0.645	0.343	0.184	0.111	0.996
Randomiz.	0.647	0.647	0.418	0.218	0.168	0.996
			Logs			
State EFI	0.008	0.018	-0.001	0.031	0.050	0.014
	(0.011)	(0.015)	(0.021)	$(0.017)^{*}$	$(0.025)^{**}$	(0.027)
Robust	0.451	0.145	0.976	0.040	0.031	0.578
Clustered	0.528	0.293	0.979	0.121	0.084	0.658
Bootstrap	0.478	0.231	0.976	0.077	0.050	0.617
Randomiz.	0.599	0.191	0.981	0.056	0.048	0.665
N	504	504	504	504	504	504
States	42	42	42	42	42	42
Year dummies	yes	yes	yes	yes	yes	yes
State dummies	yes	yes	yes	yes	yes	yes
State trends	yes	yes	yes	yes	yes	yes

Notes: The analysis is for the period 1988-1999. The dependent variables measures the number of filers in a certain income group and is measured in either shares or logs. 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 bootstrap procedure. In addition, the table also presents p-values based on robust standard errors, clustered standard errors (at the state level) as well as the randomization test. $***$, $**$ and $*$ indicate statistical significance at the 1, 5 and 10 percent levels, respectively.

Table 9: Understanding the EITC response — Marital Status

	Returns	People	Joint	Single	Head	Joint EITC	Single EITC	Head EITC
				Shares, 10-30K				
State EFI	0.616	0.671	0.055	0.037	0.464	0.295	0.011	0.595
	(0.468)	(0.564)	(0.247)	(0.366)	(0.326)	(0.172)*	(0.052)	(0.324)*
Robust	0.157	0.240	0.805	0.921	0.052	0.056	0.844	0.011
Clustered	0.243	0.293	0.844	0.927	0.211	0.126	0.846	0.111
Bootstrap	0.184	0.232	0.823	0.920	0.157	0.089	0.826	0.064
Randomiz.	0.218	0.317	0.841	0.933	0.086	0.104	0.811	0.026
				Logs, 10-30K				
State EFI	0.031	0.031	0.014	0.012	0.058	0.063	0.032	0.092
	(0.017)*	(0.016)*	(0.030)	(0.022)	(0.046)	(0.043)	(0.033)	(0.053)*
Robust	0.040	0.034	0.534	0.600	0.112	0.087	0.386	0.027
Clustered	0.121	0.097	0.670	0.630	0.263	0.190	0.392	0.125
Bootstrap	0.077	0.057	0.639	0.593	0.207	0.140	0.327	0.079
Randomiz.	0.056	0.054	0.625	0.660	0.150	0.135	0.339	0.051
N	504	504	504	504	504	504	504	504
States	42	42	42	42	42	42	42	42
Year dummies	yes	yes	yes	yes	yes	yes	yes	yes
State dummies	yes	yes	yes	yes	yes	yes	yes	yes
State trends	yes	yes	yes	yes	yes	yes	yes	yes

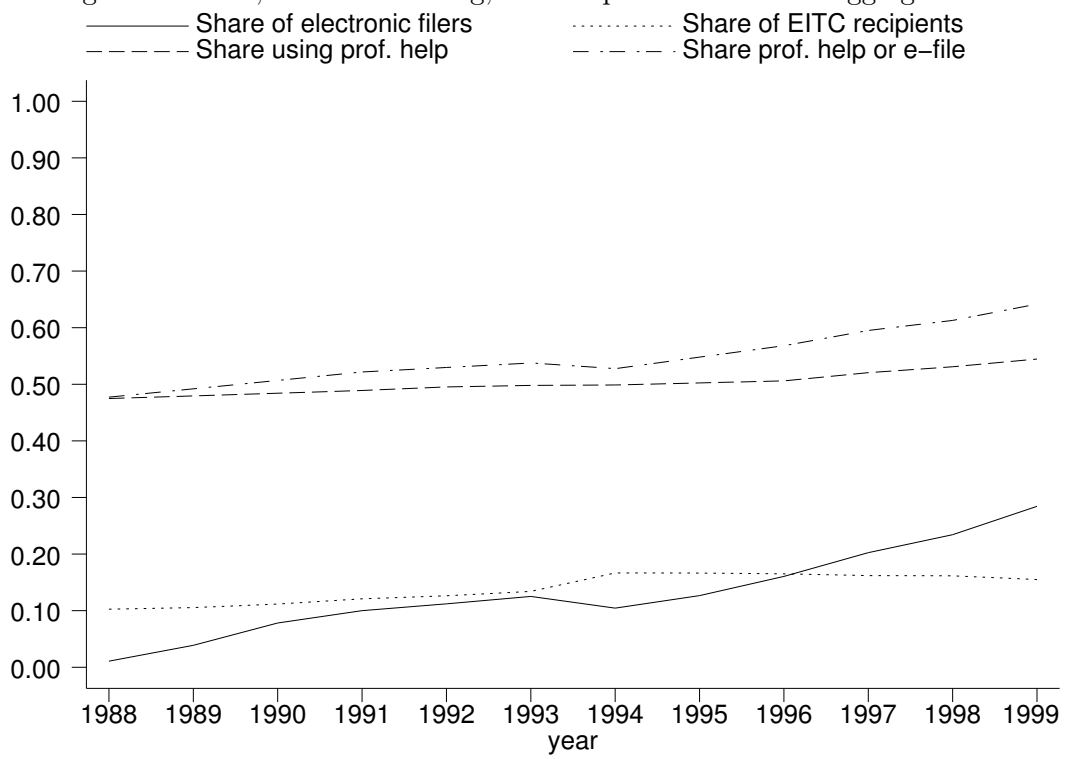
Notes: The analysis is for the period 1988-1999. The dependent variable measures the number of filers in the 10-30K category in a particular filing status category. The variables are measured in either shares or logs. All regressions include year and state fixed effects and state specific time trends. State EFI refers to the average treatment effect. Standard errors are in parentheses and are based on the standard bootstrap procedure. In addition, the table also presents p-values based on robust standard errors, clustered standard errors (at the state level) as well as the randomization test. $***$, $**$ and $*$ indicate statistical significance at the 1, 5 and 10 percent levels, respectively.

Table 10: Understanding the EITC response — Children

	Number of children			EITC taxpayers with		
	Exempt	EITC	No EITC	No	One	Two
			Shares, 10-30K			
State EFI	0.764	1.132	-0.368	0.067	0.556	0.279
	(0.683)	(0.653)*	(0.278)	(0.055)	(0.221)**	(0.214)
Robust	0.188	0.037	0.214	0.210	0.007	0.147
Clustered	0.326	0.133	0.242	0.284	0.031	0.251
Bootstrap	0.262	0.081	0.189	0.232	0.012	0.186
Randomiz.	0.260	0.066	0.278	0.341	0.012	0.216
			Logs, 10-30K			
State EFI	0.061	0.105	-0.048	0.048	0.128	0.064
	(0.042)	(0.051)**	(0.045)	(0.037)	(0.048)***	(0.046)
Robust	0.085	0.013	0.388	0.176	0.004	0.104
Clustered	0.205	0.073	0.348	0.249	0.022	0.223
Bootstrap	0.151	0.040	0.288	0.192	0.009	0.165
Randomiz.	0.118	0.036	0.425	0.278	0.009	0.170
N	504	504	504	504	504	504
States	42	42	42	42	42	42
Year dummies	yes	yes	yes	yes	yes	yes
State dummies	yes	yes	yes	yes	yes	yes
State trends	yes	yes	yes	yes	yes	yes

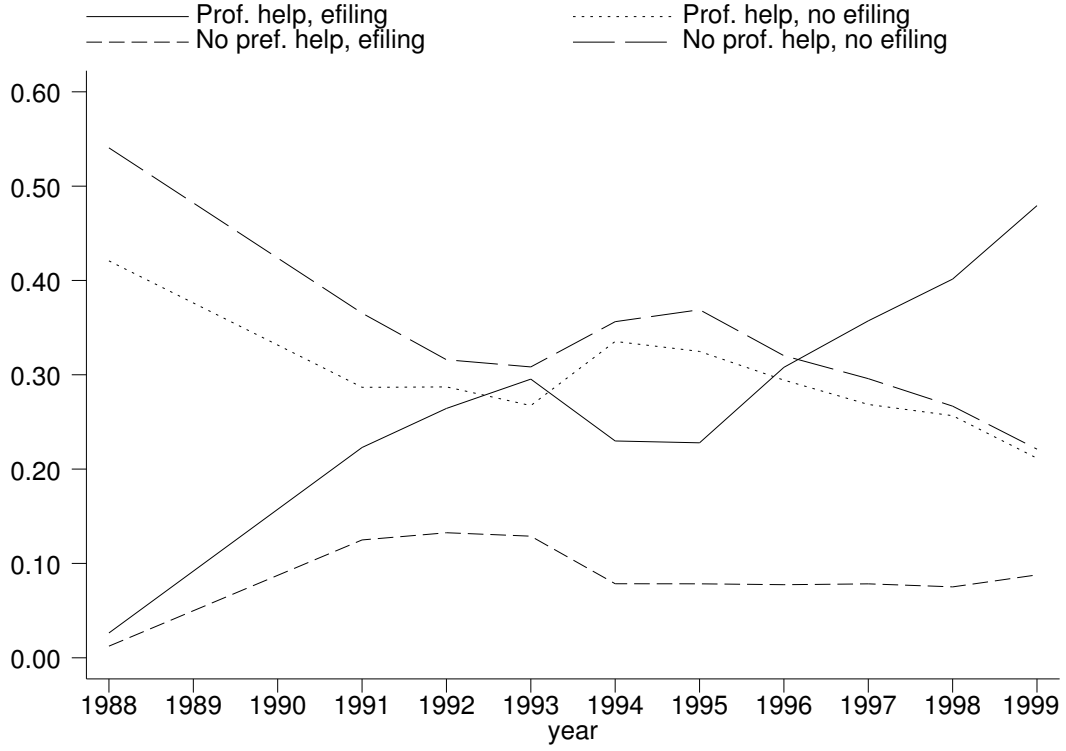
Notes: The analysis is for the period 1988-1999. In the first three columns the dependent variable measures the number of child exemptions in the 10-30K category (and also by EITC status). In the last three columns, the dependent variable is the number of EITC taxpayers with no, one and two children, respectively. The variables are measured in either shares or logs. 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 bootstrap procedure. In addition, the table also presents p-values based on robust standard errors, clustered standard errors (at the state level) as well as the randomization test. $***$, $**$ and $*$ indicate statistical significance at the 1, 5 and 10 percent levels, respectively.

Figure 1: EITC, Electronic Filing, Self-Prepared Returns — aggregate



Source: Author's calculations based on the Tax Model Files issued by the Statistics of Income division of the IRS.

Figure 2: Electronic Filing and Professionally Prepared Returns, EITC recipients, 10-30K

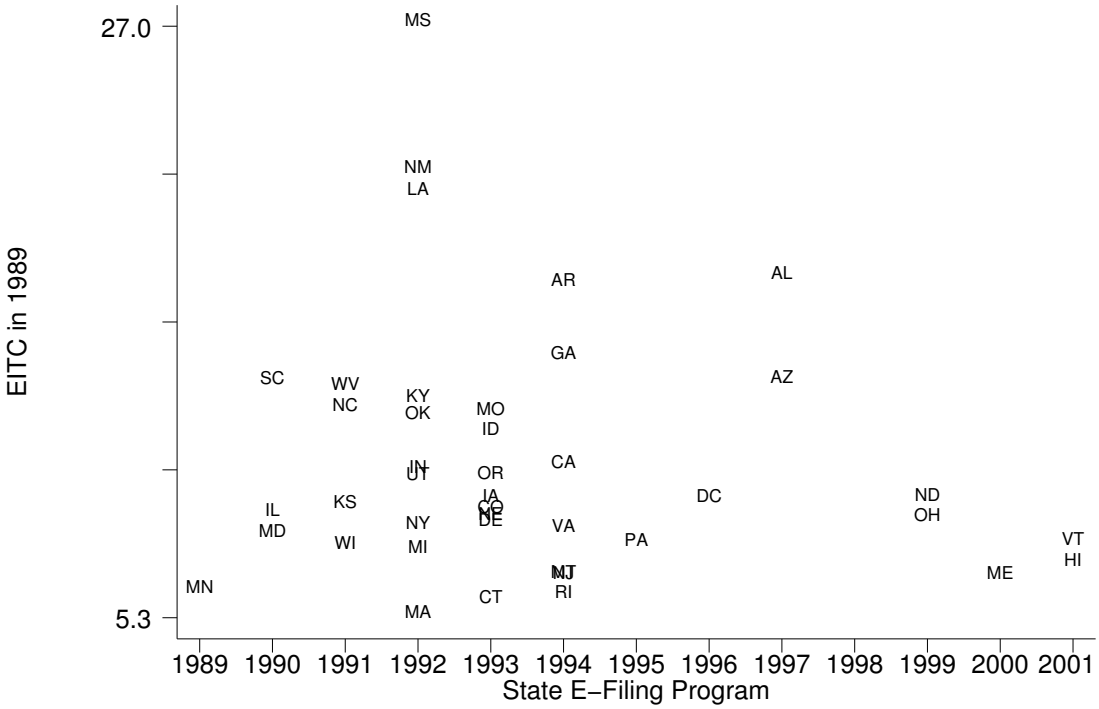


Source: Author's calculations based on the Tax Model Files issued by the Statistics of Income division of the IRS.

Figure 3: Non-participation rate

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

Figure 4: EITC in 1989



Source: The date of the introduction of a state e-filing program is taken from Table 1c. 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.

Figure 5a: Electronic Filing, shares

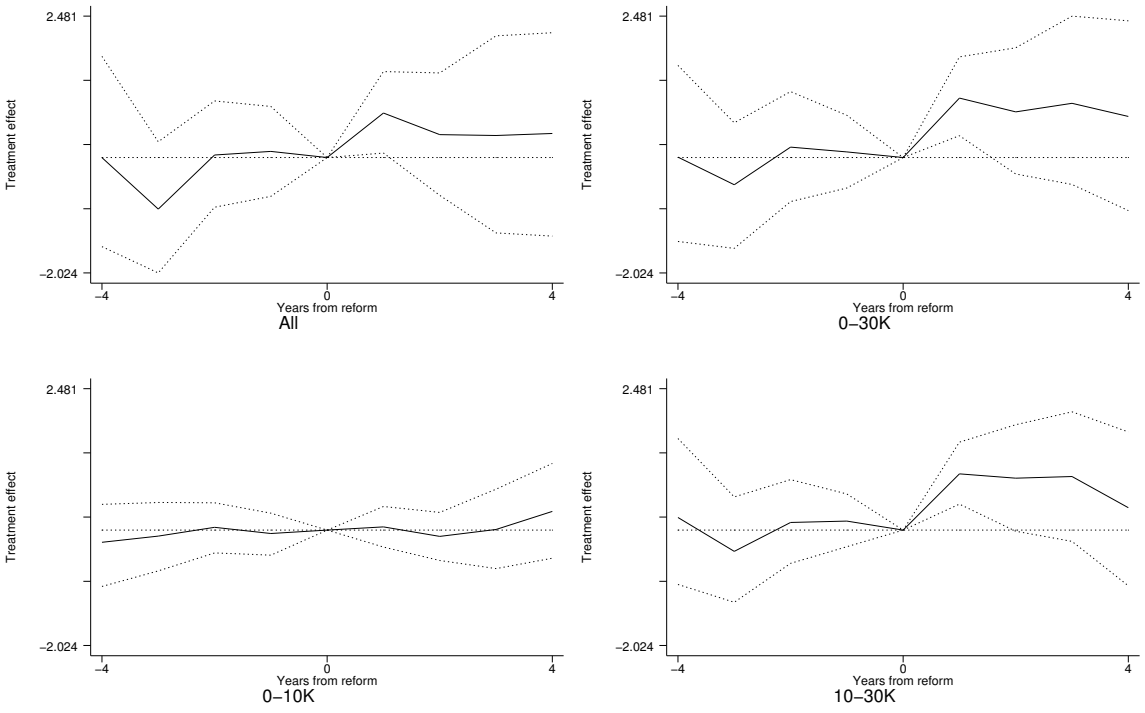
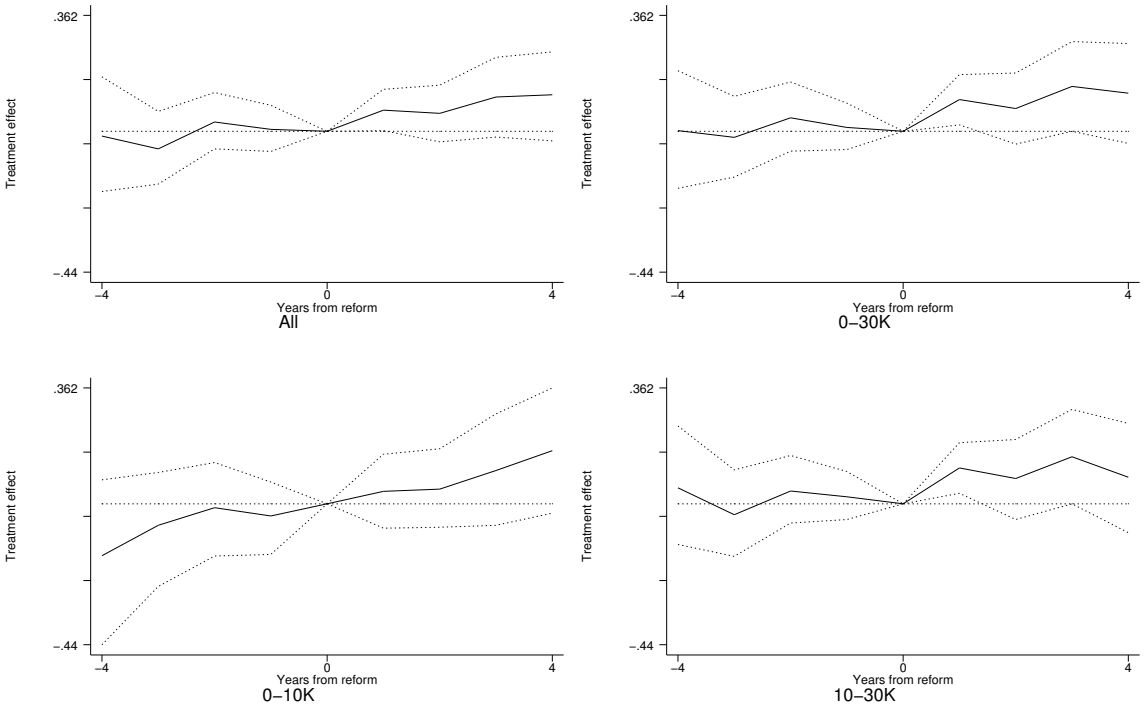


Figure 5b: Electronic Filing, logs



Notes: The figures show the coefficients from a version of equation 2 that includes both leads and lags, as well as state-specific trends. The figures also show the 95% confidence intervals constructed using the bootstrap technique (pointwise).

Figure 6a: EITC, shares

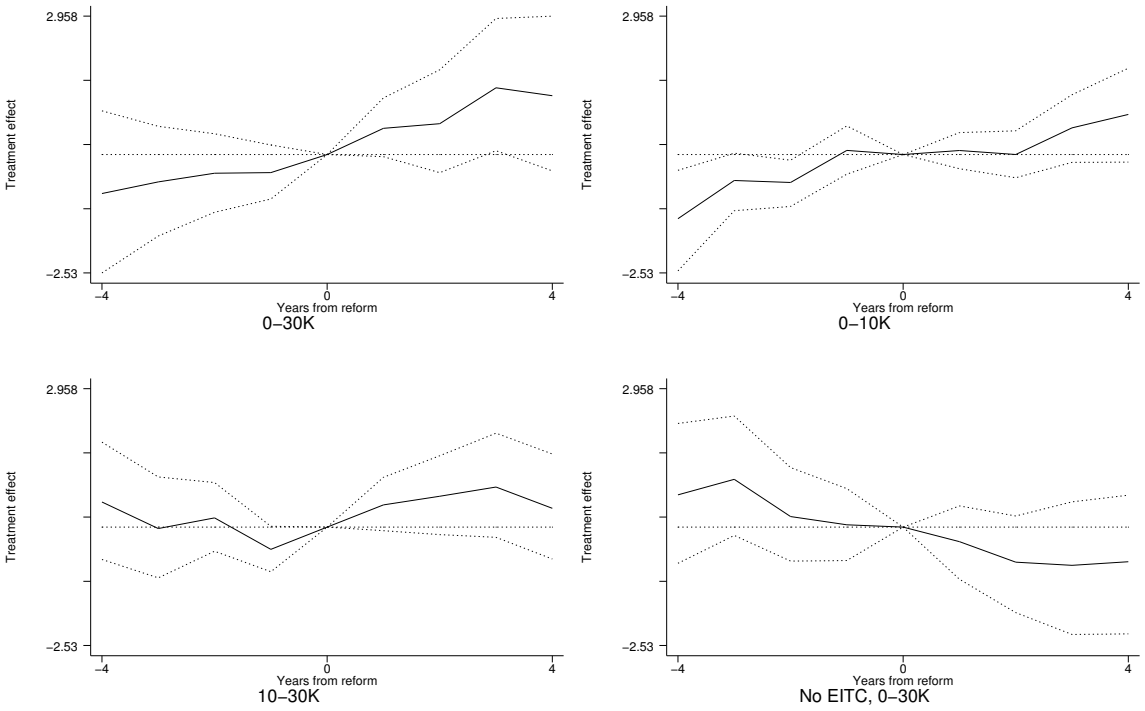
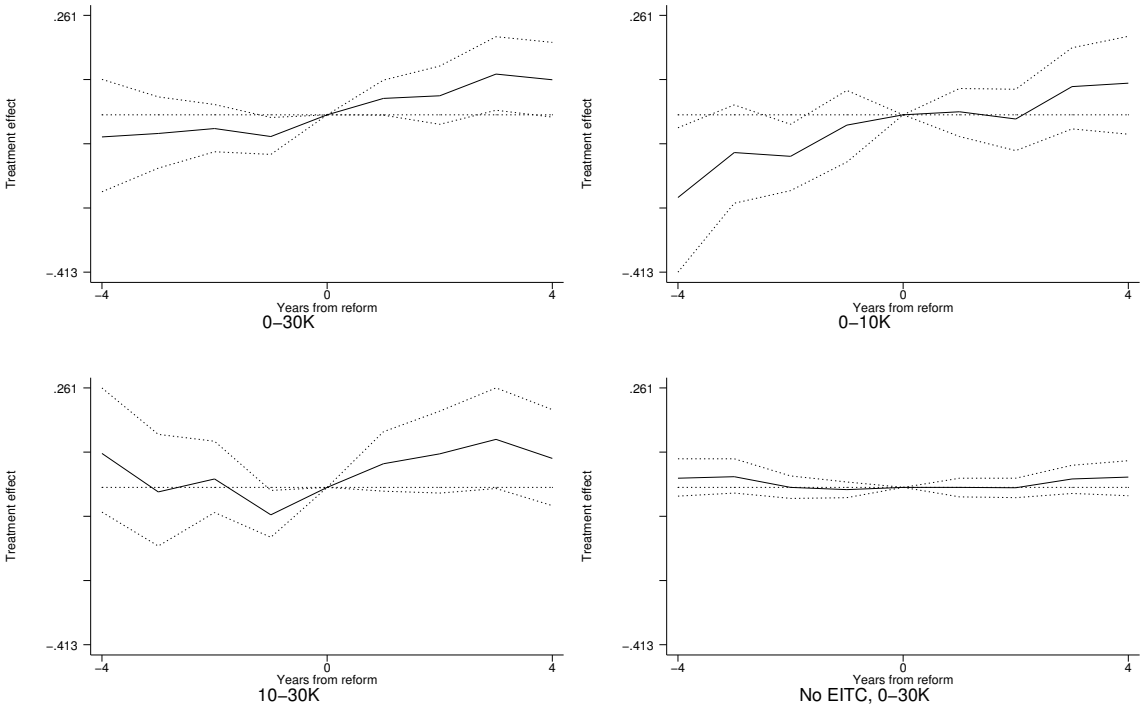


Figure 6b: EITC, logs



Notes: The figures show the coefficients from a version of equation 2 that includes both leads and lags, as well as state-specific trends. The figures also show the 90% confidence intervals constructed using the bootstrap technique (pointwise).