The Impact of Social Networks on EITC Claiming Behavior*

Riley Wilson

December 3, 2019

Click here for latest version

Abstract

Social networks have the potential to affect labor market decisions and program participation. Using the Social Connectedness Index (Bailey et al., 2018b) to capture county-to-county Facebook linkages, I explore what happens to county-level Earned Income Tax Credit (EITC) claiming behavior when the county's out-of-state social network is exposed to a newly implemented state EITC. When the number of out-of-state friends exposed to a state EITC increases the composition of EITC claims shifts toward more EITC households claiming self-employment income. The income distribution of EITC claiming households also shifts, moving away from the tails of the EITC region with smaller credits, towards the income levels that generates the largest EITC credit. This mimics the direct impacts of state level EITC policies on filing behavior, consistent with social networks providing information or increasing salience about EITC policy.

Keywords: program participation, social networks, EITC, self employment
JEL Codes:

*Brigham Young University, Department of Economics, 435 Crabtree Technology Building, Provo UT, 84602. Email: riley_wilson@byu.edu. I am grateful to Robert Moffitt, Heath Witzen, Cody Tuttle, and seminar participants at WEAI and Brigham Young University for useful comments and discussion. Clayson Shumway provided excellent research assistance in collecting data from Twitter.

1 Introduction

There is a longstanding interest in understanding how social networks affect economic decisions and program participation. For data reasons, most of the literature has focused on geographically proximate networks or familial networks. However, a person's social network might be much broader including acquaintances, firms, or other connections in more distant places that a person maintains through technology. These online and geographically distant social networks might also affect individuals' and households' decisions. To shed light on this relationship, I explore the impact of distant social networks on tax filing behavior. Using the Social Connectedness Index measure of county-to-county Facebook linkages (Bailey et al., 2018b) to capture the geographic spread of each county's social network, I estimate what happens to county-level EITC claiming behavior when a county's out-of-state social network is exposed to a newly implemented state EITC.

As states implement EITC policies, I trace out changes in the number of out-of-state friends that face a state-level EITC and see how this affects EITC filing behavior using county-level EITC filing data between 2000 and 2013. To do this I compare counties with relatively more Facebook friends in EITC expansion states to similar counties in the same state that had relatively fewer Facebook friends exposed to state EITCs before and after the policies come into place. For example, Prince George's County Maryland, has approximately 3.5 Facebook friends in North Carolina per person, while neighboring Montgomery County Maryland only has 1.6 friends per person. However, both of these counties have approximately one Facebook friend in Illinois per person. My estimation strategy compares places like Prince George's County to places like Montgomery County to see if there are differential changes in EITC claiming when North Carolina introduced a state EITC but not when Illinois did.

Even though theses policy changes have no direct effect on these households, I find that exposure to the policy through the social network changes filing behavior. The share of tax returns claiming the EITC is not affected by social network exposure, but the composition of EITC claims changes. There is an increase in EITC claims with self-employment income (filed IRS Schedule C, E, or F) and a reduction in claims without self-employment. A one standard deviation increase in the number of out-of-state friends exposed to a state EITC increases the share of returns claiming the EITC with self employment by 0.36 percentage points (8.7 percent). There is also a shift in the income distribution of EITC filers, with households claiming income that moves them away from income levels near the tails of the EITC schedule and towards the income level that generates the largest EITC credit. Trends in filing behavior are similar in high and low treatment areas prior to state expansions, and this pattern of results is robust to various specifications and sample restrictions.

These results are consistent with informed households reporting more self-employment earnings in ways that increase their potential EITC credit. This is a pattern that has been documented in previous work highlighting the bunching of self-employed workers at EITC maximizing income points (Saez (2010); Mortenson and Whitten (2018)). Upon further examination, this behavior mimics the direct impacts of state-level EITC policies on filing behavior in the state where the EITC is expanded, consistent with social networks providing information or increasing salience about EITC policy. This would suggest that even distant networks can affect households' economic decisions. Consistent with distant, online networks contributing to the spread of information, I find that at the national level, 26 percent of EITC related tweets are re-tweeted and 33 percent are "liked", with the average tweet receiving one re-tweet and 1.8 likes. Also at the local level, during tax season Google search interest for the EITC and self employment related terms increases when the out-of-state social network becomes more exposed to state EITC policies.

There is a large literature exploring the impact of social networks and peers on individuals' interactions with public programs.[1] For data reasons, most of this work has used geographic proximity, ethnic and racial background, and family relationship to construct measures of

[1] There is also a parallel literature exploring the impact of social networks on labor market decisions and outcomes (Gee et al. (2017); Hellerstein et al. (2011)).

social networks. The earliest work combined close geographic measures (e.g., zip code, public use micro area, or metropolitan area) with race, language, and ethnicity identifiers to explore how social networks affect individuals' take-up of cash welfare (Bertrand et al., 2000), public prenatal care (Aizer and Currie, 2004), and program participation among immigrants (Borjas and Hilton, 1996). Identifying the causal effect of social networks on participation decisions is complicated by the endogenous formation of networks, reflection problems, and the common or correlated unobservable characteristics among network members (Manski (1993); Dahl et al. (2014b)). For example, households in the same zip code might have other common characteristics (e.g., income, education) or common shocks (e.g., local labor market downturn) that affect participation or the outcome of interest. This might still be the case when refining the social network measure to individuals of the same racial or ethnic group within a zip code if the group identifier is also correlated with other observable or unobservable measures, like education, income, occupational choice, or preferences.

More recent work has taken advantage of rich administrative data to construct family and co-worker networks. Data from Norway have been used to show that new fathers' use of paternity leave is affected by brothers' or co-workers' interactions with the program (Dahl et al., 2014b), childrens' use of public disability insurance is affected by parental receipt of disability insurance (Dahl et al., 2014a), social insurance use is affected by neighborhoods, families, and former schoolmates (Markussen and Roed, 2015), and tax avoidance behavior is affected by the behavior of family members (Alstadsaeter et al., 2018). In many cases, these examples combine quasi-experimental variation in who is treated by a policy change, and then look at how this affects the decisions of a pre-existing network. However, to some extent these measures of networks might face the same concern of common shocks. For example, co-workers might face a similar change at the workplace that affects program participation decisions.

There are some examples in the US, but often at a smaller scale. For example, Duflo and Saez (2003) construct a field experiment at one university to see how co-workers' saving behavior changes after one member of the group is randomly nudged to attend a retirement enrollment meeting. Such detailed network data are less common or universal in the United States. However, this has recently changed with the availability of social media data. Some researchers have collaborated with Facebook personnel and used de-identified microdata from Facebook to look at how social networks affect things like job finding (Gee et al., 2017) or housing decisions (Bailey et al., 2018a). Bailey et al. (2018b) have taken this one step further by aggregating up de-identified Facebook microdata to construct county-to-county level measures of Facebook friend linkages. This measure is known as the Social Connectedness Index (SCI) and allows researchers to capture a measure of county-level online social networks.[2]

This paper builds on this emerging literature to see how distant social networks affect households' tax and earned income tax credit (EITC) filing decisions at the county-level. Thus far the literature is silent on how much distant networks matter for program participation more generally, and low-income tax policy in particular. In recent years, the EITC has become one of the largest components of the US safety net. Understanding how social networks affect EITC claiming behavior in particular would be difficult if geographic proximity was used to measure social networks, because much of the policy variation (state expansions) is geographic specific. By exploiting out-of-state social networks in this context, I am able to estimate how people's behavior changes when their out-of-state friends are exposed to a policy but they are not. As such I can separate the social network effect from the direct effect of the policy change. This context differs slightly from the previous work as the policy people face does not change, only the policy exposure of the social network. These results shed light on the important role of distant social networks in labor market and program participation decisions and suggest they should not be overlooked.

[2]For details on how to access this data, please see (Bailey et al., 2018b).

2 The EITC and EITC Claiming

The EITC is a refundable federal tax credit with explicit work incentives provided to low-income households. It has also become one of the largest anti-poverty tools in the US. In 2017 27.4 million returns received the EITC, for a total of 66.7 billion dollars (IRS, 2018). The size of the credit depends on a household's earned income. A household with zero dollars of earned income gets zero dollars of credit. As earned income increases, the credit increases until it eventually plateaus and then is gradually reduced. The phase-in rate, phase-out rate, and maximum credit varies by the number of eligible children in the household. For example, for a household with two children the credit increases at a rate of $0.40 on the dollar, peaking at a maximum credit of $5,716 (2018$) at an earned income of $14,290, and then begins to be taxed away at $0.21 on the dollar for earnings over $18,660.

Estimates of EITC participation conditional on eligibility range between 70-90 percent (Nichols and Rothstein, 2016). This is much higher takeup than observed in other transfer programs. One reason takeup of the EITC might be high, is its placement in the tax code. As such, even households who were not previously aware of the policy can qualify and apply ex post, as long as they have earned income (wage earnings or self-employment income) they can report.

A large literature has explored how the EITC has affected household decisions. The overall consensus is that the EITC has had negligible impacts on marriage and fertility decisions, positive effects on children's education and long run outcomes, strong extensive margin labor supply effects (elasticity around 0.7), and inconclusive effects on the intensive margin of labor supply (see Nichols and Rothstein (2016) for a comprehensive review of the literature examining the impact of the EITC).

There are several potential reasons the previous evidence on the intensive margin is mixed. First, repeated cross-sectional data are not well suited for identifying intensive margin effects, when there is also selection at the extensive margin. Work exploiting within person changes in EITC generosity have tended to find positive intensive margin effects (Chetty et al. (2013); Wilson (2018)). Second, individuals might lack information or knowledge about the marginal incentives. In a randomized field experiment where tax preparers explain the marginal incentives of the EITC to recipients, Chetty and Saez (2013) find that providing more information does not change individuals' labor supply on average. However, there is heterogeneity across tax preparers with some having significant impacts. Chetty et al. (2013) exploit within person changes in EITC generosity due to moves and births to see what happens to earnings at the intensive margin. They find that in areas with more "knowledge" of the program, there is a shifting in the earnings distribution, consistent with people learning and adjusting labor supply in response. Finally, individuals might face labor market frictions that keep them from freely adjusting at the intensive margin. This seems particularly relevant for low-income wage workers, where they might have little power to adjust things like their weekly hours worked. Consistent with this hypothesis, Saez (2010) shows that among wage workers there is no evidence of bunching at the EITC kink points where the marginal tax rate changes. However, among self employed workers there is substantial bunching in earnings at the first kink point in the EITC schedule, consistent with self-employed workers having more choice over intensive margin quantities. Saez (2010) suggests the observed behavior is also consistent with tax evasion.

The differences in EITC behavior between wage workers and the self employed has been of interest also. Not only does the earnings distribution of self-employed workers bunch at the first kink of the EITC schedule, but self employment income also increases after the birth of a child in high EITC "knowledge" areas (Chetty et al., 2013). The incidence of self-employment income claiming also increases after expansions in the EITC (Lalumia, 2009). Because self-employment earnings are not also reported by a second entity (such as an employer provided w-2), there is more flexibility over what gets counted as self-employment income, and people might start reporting things like babysitting if it improves their tax situation (Nichols and Rothstein, 2016). For this reason, it is important to understand how the EITC affects both wage earning and self-employment income behavior and reporting.

Although there is no prior work examining the impact of social networks on EITC claiming behavior directly, some of the previous work is closely related. The most closely related work by Chetty et al. (2013) use differences in bunching at the first kink point to proxy for differences in local knowledge of the program and find that when people move to areas with more bunching or "knowledge" of the EITC, their income changes in a way that increases the EITC refund. This would be consistent with local networks providing information or assistance in the filing process. Additionally, although Chetty and Saez (2013) do not find evidence that information provision by tax preparers affects earnings on average, they do find heterogeneous effects, with some tax preparers having large impacts. This is also what we would expect if networks influenced filing behavior. By exploiting variation in Facebook linkages, I will be exploiting a different component of the social network, one not defined by local geographic proximity. This can help us determine if distant networks have a direct effect on claiming behavior.

The previous literature exploring behavioral responses to the EITC has focused on two sources of variation: (1) variation generated by expansions in the federal credit that affect families differently according to the number of children, or (2) state-level implementations and expansions. In the last 20 years there has only been one federal expansion in 2009. This affected a small subset of the population by increasing generosity for families with three or more children. However, since 1999, 17 states have implemented supplemental state EITC policies that typically pay out an additional percentage of the federal EITC (see Figure 1). These policies vary in generosity, from 3.5 percent to 50 percent or more. Often these state-level expansions are accompanied by an increase in funding for outreach and educating the public about the tax credit, making this a plausible setting to explore the effects of social networks.[3] As states introduce and advertise EITC expansions, local residents are likely to

[3]For example, in the first two years after the California expansion $2 million was budgets for educational and outreach grants to third parties (https://lao.ca.gov/Publications/Report/3826). The National Conference of State Legislatures also reports that Iowa, Maine, Oregon, Vermont and Virginia require outreach while Iowa, Oklahoma, and Virginia appropriate funds to facilitate filing among

become more familiar with the EITC and there is some evidence that state EITC expansions increase federal EITC claiming (Neumark and Williams, 2019). In turn, they might become more likely to share this information and familiarity with those in their social network, either through liking, linking, and re-tweeting EITC related content; posting or sharing personal content on online media; or through other forms of communication (email, phone, or face-to-face). Using Facebook friendship linkages to capture the social network does not mean the content must be shared through social media, but rather proxies for geographic linkages in overall communication patterns.

3 Methodology

If people in your social network become exposed to the EITC or more information about the EITC, this can change your own awareness and knowledge of the policy. However, these effects cannot be identified if the individual's own direct exposure to the EITC is also affected. For example, if your state of residence introduces a state-level EITC, the change in policy or awareness of the policy might have a direct effect on your EITC claiming behavior, but the additional exposure and awareness of your local social network might also indirectly affect your claiming behavior. For this reason I focus on how your EITC claiming behavior changes when people in your out-of-state network face a change in state EITC policy. These acquaintances face new incentives and become potentially more aware, but your tax filing incentives remain the same. By combining variation in state-level EITC expansions with county-level geographic variation in Facebook friendship networks, I can estimate how tax filing behavior in a county changes when their out-of-state network is quasi-randomly more exposed to the EITC.

EITC-eligible families (http://www.ncsl.org/research/labor-and-employment/earned-income-tax-credits-for-working-families.aspx).

3.1 Identifying the Direct Effects of State EITC Implementation

Before examining how state EITC implementation affects filing through social networks, I first explore the direct impact on EITC filing behavior in the state that implement the EITC. If networks truly play a role, we would expect to see people in counties socially linked to the EITC-expanding state to adjust in ways similar to their Facebook friends directly affected by the change. To estimate these direct impacts, it seems natural to consider a generalized fixed effects approach that has been used in the past to explore the impacts of state EITCs on other outcomes,[4] as follows

$$Y_{ct} = \theta any\ State\ EITC_{st} + X'_{ct}\Gamma + \phi_c + \delta_t + \varepsilon_{ct}. \tag{1}$$

By examining outcomes like the EITC filing rate (percent of tax returns in the county that claimed the EITC) or the percent of tax returns that claim the EITC and additionally do (or do not) claim self employment income (as proxied by filing a Schedule C, E, or F), I can observe how both the levels and composition of EITC filing changes. *Any State EITC* is a binary variable, so the coefficient θ will be the effect on the outcome associated with having a state EITC.

The identifying assumption in this generalized fixed effects framework is that EITC filing behavior in counties in expansion states would have evolved similarly to filing behavior in non-expansion counties if the states had not implemented an EITC policy. This strategy also assumes that the stable unit treatment value assumption (SUTVA) must hold (Rubin, 1986), namely, that these state policies don't affect households in other states or counties. If in fact these expansions affect EITC filing behavior of the social network, this assumption is violated. If counties in the social network also respond, but are in the counterfactual, the estimates from equation (1) would be biased downward.

[4]This strategy has been used in the past to explore the impact of the EITC on earnings and poverty status (Neumark and Wascher, 2001), wages (Leigh, 2010), health outcomes (Baughman and Duchovny, 2013), children's educational outcomes (Bastian and Michelmore, 2018), and self-employment (Michelmore and Lim, 2018).

For this reason I will adopt an alternative approach that does not compare "treated" counties to counties that might also be responding through the social network. Instead, I rely on *within* state county-level variation in the fraction of households that claimed the EITC in 1999 (the beginning of the sample period) as follows

$$Y_{ct} = \theta(\textit{any State EITC})_{st} * (\textit{EITC Claiming Rate in 1999})_c + X_{ct}'\Gamma + \phi_c + \delta_{st} + \varepsilon_{ct}. \quad (2)$$

Here the explanatory variable of interest is the interaction between the indicator that equals one if the state has an EITC and the percent of households in 1999 that claimed the EITC. This share captures the fraction of the population that was exposed to the EITC at the beginning of the sample. Plausibly, counties with more households near the EITC range would have more scope to respond. The coefficient θ is the effect on filing behavior associated with a one percentage point increase in the 1999 EITC filing rate when the state has an EITC in place. A vector of time varying county-level gender and race shares is included, as well as county and state by year fixed effects. Standard errors are corrected for clustering at the state level and observations are weighted by the county population in 2000.

The state-by-year fixed effect makes this a comparison between counties in the same state and year. As such, equation (2) estimates how EITC filing behavior changes in counties with more potentially exposed households relative to counties in the same state, but with fewer potentially exposed households after the state EITC is implemented. These estimates would be biased if the county's 1999 EITC claiming rate is correlated with other characteristics that also change EITC filing behavior when a state EITC is implemented. I will report the estimates from equation (2) which relies on the within state variation, but also provide the estimates from equation (1) in Appendix Tables A1 and A2. I also examine responses to two similar measures that exploit within state variation. First, I construct the percent of total returns in 2000 with adjusted gross income below \$40K at the county-level as an alternative way to capture a county's ability or likelihood of responding. Second, I construct a measure of "bunching" near the EITC kink point in 2000 to capture variation in the county's knowledge about the EITC (Chetty et al., 2013) by calculating the percent of EITC claiming households with income between 5 and 15 thousand in 2000. I use the level from 2000 because the data to construct these two measures only becomes available in 2000. In Appendix Table A3 and A4 I report the results from regression equation (2) using these measures rather than the 1999 EITC claiming rate.

3.2 Identifying the Effects of Social Network Exposure

Once I have documented how state expansions affect EITC filing behavior in the expanding state, I will look to see how behavior in the social network is affected. Using the Social Connectedness Index (SCI) from Bailey et al. (2018a), I observe a scaled measure of the number of Facebook friends that active Facebook users in county c have in every other county in the country. As seen in Figures 2 and A1, some counties have few out-of-state connections that become exposed to state EITCs, while others have large, geographically dispersed networks where more connections become exposed to state EITCs. This is even true within state. Combining this network measure with state-level EITC expansions allows me to measure changes in the out-of-state social network's exposure to the EITC for each county as follows

$$Network\ Exposure_{ct} = \frac{1}{pop_{c,2010}} \sum_{j \neq s}^{S} \gamma(friends\ in\ state\ j)_c * 1(State\ EITC\ in\ j)_t. \quad (3)$$

In other words, I sum up the total number of friendship links for county c in states s that currently have a state EITC in place. However, I exclude the state that county c is in (state s) to only look at out of state friendship links. This number is then divided by the population in county c in 2010 to identify the number of out-of-state friends per person that are exposed to a state EITC. Notice that this measure will change over time as states implement EITCs, and counties will be more or less exposed depending on the number of friendship links they have. I have also included a parameter γ in equation (3). This is because the Social Connectedness Index reports county-to-county friendship links at a scalar multiple of the true value, for privacy. In other words, I do not observe $(\textit{friends in state } s)_c$ but $\gamma(\textit{friends in state } s)_c$. As such, the units of this measure are not inherently meaningful. For this reason I standardize network exposure by subtracting the sample mean, and dividing by the standard deviation. As such, a one unit increase is now equivalent to a one standard deviation increase in the number of out-of-state friends per person exposed to a state EITC. In regression tables, I will refer to the standardized *Network Exposure* as the "number of Out-of-State Friends per person Exposed to a State EITC".

I estimate the impact of network exposure to the EITC on EITC filing behavior as follows

$$Y_{ct} = \beta \textit{Network Exposure}_{ct} + X'_{ct}\Gamma + \phi_c + \delta_{st} + \varepsilon_{ct} \tag{4}$$

I examine the same outcomes as above: the EITC filing rate, and the percent of tax returns that claim the EITC and do or (do not) claim self employment income (as proxied by filing a Schedule C, E, or F). The main explanatory variable is the number of out-of-state friends per person exposed to a state-level EITC. Because this is standardized, the coefficient β can be interpreted as the change in EITC filing rates associated with a one standard deviation increase in the number of out-of-state friends exposed to a state level EITC. As such it is not directly comparable in magnitude to the coefficient θ from the model above. County-level fixed effects are included to control for time invariant county characteristics. State-by-year fixed effects control for state-level changes over time, effectively making this a comparison between counties in the same state. As such, any variation over time at the state-level (including if ones own state expands the EITC) will be absorbed. I also include a vector of time varying county-level gender and race shares ($X'_{ct}\Gamma$). Standard errors are adjusted to correct for clustering at the state-level and observations are weighted by the county population in 2000.

Because I have included state-by-year fixed effects, the specification in equation (4) is comparing counties in the same state over time. The identifying variation comes from within state differences in the geographic spread of the county-level Facebook network. In other words, I am testing to see how EITC filing behavior changes in counties with more Facebook linkages in states that expand the EITC relative to counties in the same state –but with fewer Facebook linkages in expansion states– after the state EITC is implemented. Because I am looking over all state EITC policies, the identifying assumption is that counties with more Facebook linkages in expansion states would have behaved like other counties in the same state that had fewer Facebook linkages in expansion states if the state EITCs would not have implemented.

In Columns (1) and (2) of Table 1 we see that there are level differences at the beginning of the sample period between counties that experienced larger increases in the number of out-of-state friends exposed to state EITCs and counties that experienced smaller increases. Counties that experienced larger increases had more minorities, were more educated, and had stronger labor markets. Because this strategy exploits changes over time, these level differences are not inherently problematic for identification. In Columns (3) and (4) counties that experienced both large and small increases in the number of out-of-state friends exposed to state EITCs saw similar trends in demographic characteristics. The gender and racial composition as well as the unemployment rate evolved the same on average between these two groups. However, there were slight differences in the evolution of the education composition and average earnings. The population became more educated in areas that saw larger increases in the number of friends exposed to state EITCs, and larger increases in earnings.[5]

The identifying assumption would be violated if counties with many Facebook friends in EITC expansion states were systematically different than counties in the same state with fewer Facebook friends in expansion states *and* if these differences affected filing behavior

[5]The change in earnings as a percentage is similar. It is also possible that this is an endogenous response to the policy.

over time. For example, if counties with more Facebook friends in expansion states are becoming more educated and wealthy, they might also become more savvy or engage in more EITC outreach over time, which could impact filing behavior and bias the estimates. However, since there are discrete changes in a state's EITC status, I can verify that trends in EITC filing behavior prior to the expansions are similar for counties with many and fewer friends. Parallel outcome trends prior to the treatment would rule out concerns like the example above, unless these independent activities also occur precisely when the other states expand the EITC.

3.3 Data

To estimate this relationship I combine detailed local tax filing data with the Social Connectedness Index. In a joint project, the IRS Statistics of Income (SOI) and the Brookings Institution provide annual zip code-level data on tax –and specifically EITC– filing behavior. This includes measures like the total number of returns filed, the number of returns that claimed the EITC, the total amount refunded to EITC claimants, and the number of tax returns (and EITC filing tax returns) in small income bins. It also includes counts of the number of returns that filed a Schedule C, E, or F form. Schedule C is filed for business profits or losses or sole proprietorship, and should be filed by individuals engaged in self-employment, independent contracting, or who received Form 1099-MISC. Schedule E is filed to report income or losses from real estate, royalties, partnerships, S corporations, estates, trusts, or other similar entities. Schedule F is filed to report income or losses from a farm. From the IRS SOI, in 2013, about 16 percent of returns filed a Schedule C while only 1.3 percent filed a Schedule F. Schedule E filing numbers are not provided by the SOI, but many of these events are associated with high income households, not likely to be income eligible for the EITC. As such, EITC claimants who also filed Schedule C, E, or F likely filed Schedule C for some type of self employment or independent contractor income. I then construct the EITC filing rate which is the percent of total returns claiming the EITC, as well as the percent of total returns claiming the EITC with a Schedule C, E, or F (EITC with self employment income) and the counterpart to this the percent of total returns claiming the EITC without a Schedule C, E, or F (EITC without self employment income).

This zip code level data is available from 1997 to 2014. In 1997 and 1998 only the number of tax returns, EITC claiming returns, and aggregate EITC amounts are available, and in 1999 self employment measures and income bins are not included. In 2014 only EITC specific variables are provided (so rates cannot be constructed). For this reason I use the 1999 data to construct the county level EITC claiming rate, and restrict my sample to 2000 to 2013. Because this is reported at the zip code level, I have to aggregate up zip code counts to the county level. I do this using zip code to county crosswalks provided by the US Department of Housing and Urban Development. Occasionally zip codes cross county borders, in which case I allocate the population proportionally by the share of the zip code population in each county. In practice, 75 percent of zip codes are in one county and 90 percent of zip codes have over 90 percent of their population in one county.[6]

I then link this data to the Social Connectedness Index, which was introduced and discussed in detail by Bailey et al. (2018b). The measure is based on the number of Facebook friendship links between each county and every other county in the US from a snapshot of active Facebook users in 2016.[7] The data is then normalized for privacy so that researchers observe a scalar multiple of the number of friendship links.[8] As such I observe a static measure of each county's social network, as captured by Facebook users. I link this data to Census county-level population data to measure friendship links per person.

I then link the Social Connectedness index to a panel of state-level EITC policies which were gathered from multiple sources and collected by hand. Using this data I am then able to construct the network exposure measure described above in equation (3) which I will use as

[6]The IRS SOI provide county level data, but it does not include many of the EITC related variables.

[7]An active users is "a registered Facebook user who logged in and visited Facebook through our website or a mobile device, or used our Messenger application... in the last 30 days" (Bailey et al., 2018b).

[8]Researchers interested in this data can be granted access as outlined in Bailey et al. (2018b), who have graciously provided this rich resource.

a proxy for social networks. Importantly, I view this measure as a proxy for social networks more generally, as people linked on Facebook also interact in other ways. For example, someone with Facebook friends in a given community is also likely to encounter or interact with people, firms, or the local news in that community.

4 Results

4.1 Impact of State EITC on EITC Filing in Implementing State

The direct impacts of a state EITC on filing behavior are reported in Table 2. For reference the mean EITC claiming rate in 1999 was 18 percent. Having a state EITC is associated with an insignificant 0.01 percentage point increase in the percent of returns that included an EITC claim (EITC filing rate) for every one percent increase in the 1999 EITC claiming rate. Although insignificant this would be consistent with a mean impact of 0.18 percentage point increase in the EITC filing rate.[9] In columns (2) and (3) I estimate the impact of the percent of returns that file for the EITC and whether or not they include income from self employment. Although the EITC filing rate does not significantly increase, there is a shift in the composition of filing with more people claiming self employment. The percent of returns with the EITC and no self employment goes down by a significant 0.06 percentage points, while the percent of returns with the EITC and any self employment (as proxied by filing a Schedule C, E, or F) goes up by a significant 0.07 percentage points. At a mean EITC claiming rate of 18 percent, this would suggest EITC filing with self employment increased by approximately 1.26 percentage points while filing without self employment fell by a similar amount.[10]

In Table 3 I examine the impact of state EITC implementation on the size of the EITC

[9]These impacts are in the same range as the insignificant, full population estimates of Neumark and Williams (2019) looking at how state EITC affect federal filing. They however find some evidence of significant effects among low-skilled workers with children. These estimates are also similar to estimates of state EITCs on extensive margin employment (Bastian and Jones, 2019).

[10]Pre-trends for the direct impacts of state EITCs are presented in Appendix Figure A2.

return and the income distribution of EITC filers. Having a state EITC is associated with a 6.13 dollar increase in the average federal refund for every one percent increase in the 1999 EITC claiming rate, or 110.34 dollars at the mean. This only includes EITC dollars received from the federal government, not the state. When examining the adjusted gross income distribution of EITC filers, we see a significant reduction in the share of households in the lowest bin (Under $5K) and highest bins ($35K and higher) and a significant increase in the share of households in income bins in the middle. As seen in Figure 3, these changes in the adjusted gross income distribution move households away from the edges of EITC eligibility with smaller credits towards incomes that increase the EITC credit. However, I do not observe a significant increase in the share of households in the income bins that maximize the EITC credit. This shifting towards credit increasing incomes is consistent with the patterns observed by Saez (2010) and Mortenson and Whitten (2018), which documents substantial bunching at EITC credit maximizing kink points, especially among those with self employment income. Unfortunately, the data does not allow me to examine the distribution of income for the subgroup of EITC filers that claim self employment.

The generalized fixed effect specification yields similar results (see Appendix Tables A1 and A2) although they are imprecise. EITC filing shifts towards filing with self employment, and the adjusted gross income distribution of EITC filers shifts away from the tails of EITC eligibility toward income levels that increase the EITC credit. In this specification the largest coefficients are at income levels that maximize the EITC credit. These estimates are less precise with smaller implied magnitudes, which we would expect if untreated counties in the social network also respond but are included in the counterfactual.[11]

When using the alternative measures of within state EITC variation the pattern of results is nearly the same (see appendix Tables A3 and A4). There is a shift toward reported self employment among EITC filers and movements in the income distribution away from the tails with smaller credits towards the larger credit regions. When using the percent of

[11]The pattern of results is similar if I instead use the state EITC percentage (a continuous measure) as the explanatory variable of interest.

households in 2000 with income below \$40K there is also a signficant increase in any EITC filing, suggesting there might be an increase in overall filing (as suggested by Neumark and Williams (2019)).

4.2 Impact of Network Exposure on EITC Filing Rates and Composition

Given this direct effect on the share of EITC filers reporting self employment income and the adjusted gross income distribution in states that implement state EITCs, I next report the results from equation (4) to test if filing behavior is transmitted through the social network. Remember, because I view the Facebook network as a proxy for overall social networks and connectedness, these coefficients will test whether people's connections to certain areas affect their EITC filing behavior. In Table 4 I explore the impacts on the same outcomes used in Table 2. As seen in column (1), the number of out-of-state friends exposed to a state EITC does not affect the federal EITC filing rate. However, there is a composition change in the share of returns that claim the EITC and report self employment income. When the number of out-of-state friends exposed to a state EITC increases by one standard deviation, the percent of returns that claim the EITC but no self employment falls by 0.35 percentage points (column (2)), while the percent with any self employment rises by a nearly identical 0.36 percentage points (column (3)). This results in the fraction of EITC claimants with any self employment increasing by 2.3 percentage points (10 percent). This would suggest that when households' out-of-state social network experience a state EITC implementation there is no change in the number of EITC claimants, but claimants shift toward reporting self employment income. This shifting occurs, despite the fact that the EITC incentives do not change for these people. For reference these effects are smaller than the average direct effects. The shift in self employment reporting is about the impact of an additional five percent of the population in 1999 claiming the EITC, or 28 percent of the average direct effect.

In Table 5 I estimate how the number of out-of-state friends exposed to a state EITC affects the average EITC refund and the adjusted gross income distribution of EITC filers. Exposure through the social network is associated with an insignificant increase in the average EITC refund. However, there is shifting in the income distribution. As was observed in the income distribution of EITC claimants in implementing states, households move away from the edges of EITC eligibility where the credit is smaller, towards income levels that increase, and maximize the EITC credit (see Figure 4). This shifting in the income distribution is also smaller than the average direct effect in implementing states. When households' out-of-state friends were exposed to a state EITC, EITC claimants became more likely to claim self employment income, and shift their income in ways that moved them away from smaller EITC credits towards the middle, where credits are larger. This behavior closely follows the behavior observed by those directly impacted by the state-level expansion, consistent with behaviors being discussed and passed on through social networks.

There are several ways households might adjust their reporting of self employment income in ways that would lead to a larger EITC payment. First, people might be more likely to engage in self employment or contract work. As we don't see an increase in the EITC filing rate, this would likely be people who previously claimed the EITC adding self employment. Using the American Community Survey (ACS) from 2005 to 2017, I look at how out-of-state friend exposure to state EITCs affects reported family unit level employment and self employment in Table 6.[12] I see no change in the employment rate (consistent with no change in the EITC filing rate), but a one standard deviation increase in the number of out-of-state friends exposed to a state EITC is associated with a 0.2 percentage point increase in reported self employment. This point estimate is only about half as large as the increase in EITC filing with self employment (but not statistically different), suggesting at least some of the self employment reported to the IRS actually occurred. This increase in self employment

[12]Because only large counties are identified in the ACS, I follow the method of (Autor et al., 2013) and probabilistically assign family units to commuting zones based on their public use micro area (PUMA) of residence and then collapse to get commuting zone level measures. Because I am interested in out-of-state policies I maintain people's state of residence to get commuting zone by state levels of geography. I scale the network exposure by the same standardization as above, so that the magnitude of effects can be compared.

is also concentrated among non-incorporated self employment, which is the more relevant margin for low-income households generating additional income. As seen in Appendix Table A7, this increase in self employment is also concentrated among family units with EITC eligible children and single mothers.

Alternatively, this might be a matter of reporting rather than work. As suggested by Saez (2010), households might start reporting income and expenses that were previously undocumented when they improve their position in the EITC schedule. For example, households might start reporting earnings (and expenses) from babysitting, lawn mowing, or other side jobs that would have otherwise been ignored. It is also possible that households start false reporting self employment income. In fact, the IRS has recently acknowledged the practice of filing false Schedule C income and expense claims stating, "Fictitious Schedule C's especially those with no 1099 Misc support or no supporting income or expense records that qualify for or maximize EITC is a growing problem."[13] Kuka (2014) shows that EITC expansions increase self-employment claiming but shows it appears to be driven by tax noncompliance rather than increased labor supply. Lalumia (2009) also shows that EITC expansions led to more claiming of self-employment income, particularly among those not using a paid preparer. This exposure measure is associated with a 0.50 percentage point increase in households claiming the EITC and self filing their taxes, potentially accounting for the total increase in EITC filing with self employment (see Table A8).[14] Taken together, this would suggest that the knowledge or information passed through the social network induces some additional labor force participation, but likely also increased the reporting of actual or fictitious self employment income.

[13] See https://www.eitc.irs.gov/tax-preparer-toolkit/frequently-asked-questions/earned-income-self-employment-income-and-business, accessed January 15, 2019. This admission was first pointed out by Mortenson and Whitten (2018).

[14] The rise in self filing does not necessarily indicate a rise in false reporting. Often paid preparers charge an additional fee to file a Schedule C, which might induce people to self file.

4.3 Event Study, Pre-trends

The identifying assumption in equation (4) is that counties with more Facebook linkages in expansion states would have behaved like other counties in the same state that had fewer Facebook linkages in expansion states if the states would not have implemented an EITC. This is similar to the parallel trends assumption, suggesting that these counties should have followed similar trends before the policy change occurred. However, since states expand the EITC at different times, the pre-trends are more difficult to test. Alternatively, one could think about each state-level expansion as an individual event or treatment. Then the identifying assumption is that counties with more Facebook linkages in state j would have behaved like counties in the same state, but with fewer Facebook linkages in state j if state j would not have implemented and EITC. Parallel pre-trends for each state-level event can be checked by estimating the following specification

$$Y_{ct} = \sum_{\tau=-3}^{4} \beta_\tau \big((\textit{friends in state } j)_c * 1(\tau \textit{ years from introduction})\big) + X'_{ct}\Gamma + \phi_c + \delta_{st} + \varepsilon_{ct}. \quad (5)$$

This specification includes both county fixed effects and state-by-year fixed effects, so that I am still comparing counties in the same state over time. The β_τ vector of coefficients traces out the impact of the number of friendship links county c has in state j from three years before state j introduces an EITC and 4 years after. When doing this I omit the year prior to EITC introduction ($\tau = -1$), to keep this as the reference year. As such, I can verify that counties with many friendship links in state j and counties with few friendship links in state j behaved similarly before state j introduced an EITC, as well as trace out the effect of the policy change over time.

Rather than estimate equation (5) separately for each state event and to more closely match the previous analysis, I stack the data to estimate the average affect across all implementing state events as follows

$$Y_{ct} = \sum_{\tau=-3}^{4} \beta_\tau \big((\textit{friends in state } j)_c * 1(\tau \textit{ years from introduction})\big) + X'_{ct}\Gamma + \phi_{jc} + \delta_{jst} + \varepsilon_{jct}. \quad (6)$$

The county fixed effect now becomes an event-by-county fixed effect and the state-by-year fixed effect becomes an event-by-state-by-year fixed effect to keep this a within-event comparison. As such I am still comparing outcomes between same-state counties that have strong and weak linkages to state j, but the β_τ coefficients are the average effect over all of the states j. Because states are implementing the EITC during the entire sample period, it is not possible to estimate equation (6) for all expansion states in a balanced panel. As such, I restrict the sample of expansion states to states that implemented the EITC between 2003 and 2010, so I can observe counties for the entire period (3 years prior and 4 years post). These coefficients are plotted with 95 percent confidence intervals from state clustered standard errors for the percent of returns with the EITC and self employment (in blue) and the percent of returns with the EITC and no self employment (in red) (see Figure 5).[15]

I observe no evidence of differential trends prior to the treatment. The number of out-of-state friends that will be exposed to a new state EITC does not affect the EITC filing rate –both with self employment and without self employment– before the state EITC is implemented. This would be inconsistent with counties that have more out-of-state friends in EITC expansion states being systematically different in ways that affect EITC filing behavior over time. After the state EITC is implemented the filing rate of EITC returns with and without self employment start to diverge. Starting the year of the policy, the EITC returns without self employment drops and slowly declines to about -0.6 percentage points. EITC returns with self employment income start rising after one year, and climb steadily to approximately 0.6 percentage points. As such the net effect on the filing rate is negative in the early years after treatment, and then become close to zero and noisy, as seen in the

[15] Alternatively, it might seem natural to cluster at the expansion event level. Results are robust to two-way clustering at the state and expansion event level.

previous estimates. This would suggest that EITC filing behaviors in counties with many Facebook friends in expansion state j and counties with few Facebook friends in expansion state j were on a similar trend, and only diverge after state j implemented an state EITC. As such, the results are not driven by differential demographic trends (like those in Table 1) unless they began to change precisely when the EITC was implemented.[16]

4.4 Robustness

The results presented above are consistent across a range of specifications (see Table A5). Trimming extreme values in the friendship network measure and excluding controls does not significantly change the coefficient for self employed filing. When looking at no self employment the coefficients are smaller, and less precise, but not statistically different. As we see in Table 1, there are some differences across counties, so I also include county linear trends to account for potential trend differences. The coefficient on self employment falls, but is still significant, while the coefficient on no self employment is virtually unchanged.[17] If I make this a comparison across all counties in the US by including year fixed effects (rather than state by year fixed effects) the coefficients are slightly larger in magnitude but not statistically different (0.44 and -0.49 versus 0.36 and -0.35). As distance is an important predictor of social network strength, it is possible that the social network measure is just a proxy for distance, and I am capturing the behavior of counties in the same local media market or labor market. If the state EITC affects the local media market or labor market, this could bias my estimates. To test this I include Designated Market Area (DMA) by year fixed effects, which result in slightly larger coefficients. I also exclude counties that border expansion states (and might be in the same labor market). In this case the coefficients are

[16] I also provide event study figures for the direct effects in Appendix Figure A2, and for income groups in Appendix Figure A3. When looking at the direct effects I report wild bootstrapped standard errors because there are only 11 state events. When looking at EITC filing with self employment there are no clear pre-trends but a steady rise starting in the year of the policy change. When looking at EITC filing without self employment, the drop appears to occur one year early, but the estimates are all imprecise and cannot be statistically distinguished.

[17] As noted by (Wolfers, 2006), county specific trends might pick up part of the effect of the policy, which might explain the drop in the coefficient on self employment.

larger in magnitude, but not statistically different.[18]

When the outcome is measured in logs, the coefficients have the same sign, and suggest treatment effects of a similar magnitude. The only specification that does not yield similar results is when the regressions are not weighted by the county population. However, this is unsurprising as small counties, where a few returns can vastly change the filing rates, are given more weight. If I estimate the log model (to reduce heteroskedasticity) with equal weights, the effect on self employment filing is once again positive and significant, while the effect on non self employment filing is negative, but insignificant.[19]

By exploiting the same within state variation used to estimate the direct effects, I can also relax slightly the identifying assumption. Similar to equation (3) I construct county level network measures but weight each friend county by the fraction of the returns in 1999 that claimed the EITC. This builds in an additional difference, and allows me to test if counties with a lot of friends in high EITC treatment counties in expansion states respond differently than counties with fewer friends in high EITC treatment counties. In Table A6 I find nearly identical impacts, with a shifting towards more self employment income among EITC filers.[20]

Perhaps the biggest concern is that two counties in the same state with different out-of-state social networks might be different on other dimensions as well. For example, counties with more out-of-state friends might also be more urban, which might affect EITC filing

[18]Similarly, if I examine counties in bins by the distance to the closest county that experiences an expansion, the effects remain present for counties up to 200 miles away. We might also be concern that states adopt EITC policies in geographic clusters, thus by distance the social network measure would be correlated with own adoption of state EITC. However, since I am including state by year fixed effects, I am only exploiting within state variation in the network. Nevertheless, if I regress equation (4, but include year fixed effects rather than state by year fixed effects, the coefficient on the number of out-of-state friends exposed is negative. This would suggest that if anything, states are less likely to adopt an EITC when neighboring states do, which would bias my estimates downward.

[19]There is also variation across states in the size of the EITC credit. Using these rates, I construct the average state EITC rate that out-of-state friends are exposed to (similar to equation (3). Using this measure the pattern of results is the same, but less precise. The EITC filing rate stays the same, and self employment filing as a share of EITC filing goes up.

[20]I have also interacted the network measure with whether the county had a high (above median) EITC claiming rate in 1999, to more closely correspond to the analysis of the direct effects. I find that the effects are larger in counties with more of the population in the EITC eligibility region (see Appendix Tables A9 and A10).

behavior. However, as we see in Figure 5, these potential differences must only change filing behavior precisely when out-of-state friends become exposed to a state EITC. Even though this seems unlikely, I implement an alternative approach that exploits county-to-county distance and the network of interstate freeways to isolate variation in the social network due to these longstanding characteristics. Because both distance and the freeway system pre-exist the state EITC expansions, I can use them to predict the social network, and isolate variation in the social network due to these fixed characteristics, rather than changing demographic trends which might affect the outcome. To predict the social network I estimate

$$\gamma(\textit{friends in county } j)_c = \beta_1 miles_{jc} + \beta_2 1(\textit{Same Interstate})_{cj} + \beta_3 miles_{jc} * 1(\textit{Same Interstate})_{cj} + \phi_c + \varepsilon_{cj} \tag{7}$$

The outcome is the scaled number of friends in county j of people who live in county c. This is then estimated as a function of distance (from the population centroid of county c to the population centroid of county j), an indicator for whether or not county c and county j share an interstate freeway that crosses through them (see Appendix Figure A6), the interaction of the two, and a county fixed effect. From equation (7) the predicted number of friends in county j for people in county c is created.[21] I then aggregate this up to the state level, and construct the measure analogous to equation (3), but use the predicted number of friends. I then estimate an equation similar to equation (4), but use the predicted number of out-of-state friends exposed to a state EITC. These results are presented in Table 7. The pattern of results is similar. The EITC filing rate does not change, but the composition shifts towards households filing with self employment income, with a one standard deviation increase in the predicted number of out-of-state friends exposed to a state EITC increasing the share of EITC returns with self employment by 1.3 percentage points (5.7 percent). Even when isolating variation in the social network from fixed characteristics, the results hold.[22]

[21]The coefficients from this prediction are provided in Table A11.

[22]This predicted measure is similar to a simulated instrument. If I use this measure to instrument for network exposure the pattern is the same but the coefficients are about twice as large.

Another concern is that since the Facebook network is constructed from a single snapshot in 2016, it might actually be mechanically endogenous to the outcome if people are moving from the state of treatment. For example, if people in Michigan move to other states after Michigan implemented the state EITC in 2008, they might carry their EITC filing behavior with them. They would increase the social network tie (as measured in 2016), and also make it appear as if people are responding. If people start moving from expansion states after the EITC is implemented, we should observe that the number of out-of-state friends exposed to a state EITC is predictive of in-migration. When I regress equation 4, but use the in-migration rate from states that have a state EITC as the outcome, I find a significant 0.4 percentage point increase in household migration (see Appendix Table A12).[23] To see if this increase in migration is driving the change in filing behavior, I re-estimate the effects on EITC filing and filing with self employment while controlling for in-migration rates from EITC states. The coefficients are only slightly smaller, suggesting the EITC response is not driven by a change in migration behavior.[24] Also, the previous exercise predicting the social network as a function of distance and interstate roads will eliminate changes in the social network that arise if people moved after being treated by the EITC.[25]

[23]These in-migration rates are calculated from the IRS SOI county-to-county flows. However, many of these cells are suppressed fro privacy, leading to lower overall migration when aggregating up, especially to less populated areas. If I instead look at in-migration rates from all other states (which is suppressed much less often) I get a marginally significant coefficient of 0.18.

[24]As further evidence that this is not driven by migration from EITC expansion states, employment estimates from the ACS are nearly identical when I exclude people who have moved from an EITC expansion state in the last three years.

[25]Another concern is that Facebook was not available during the entire analysis period. Although the number of Facebook friends is meant to proxy the social network, it is possible that social networks were different before and after Facebook was introduced. I split the sample into pre-2004 and post 2004 (inclusive) to see if the responses are different. It should be noted that states that implemented the EITC before 2004 and after might be different, and overall familiarity with the EITC has changed over time. I find that before 2004 increases in the Facebook social network measure are associated with increases in EITC filing and EITC filing without self employment. Only after 2004 is there a compositional shift towards filing with self employment, similar to that observed in Table 4 (see Appendix Table A13).

5 Online Networks: Twitter Words and Online Search Behavior

Although the previous evidence is consistent with households changing EITC filing behavior after more of their out-of-state social network is exposed to the EITC, it is unclear what the mechanism is for this transfer of information. The social network captured by Facebook friendship links is simply a proxy for more broad social networks, so it is not necessary that the information be transmitted directly through Facebook. Acquaintances could communicate this information through other social media, online or phone communication, or even face-to-face communication. It is also unclear what information is being passed through the social network. Members of the network exposed to the EITC might give specific information on how to maximize your refund through EITC filing, or might give general information about the potential money available through the EITC.

To understand the role of online social networks, I collect all tweets from Twitter that reference the EITC between January 2010 and September 2019.[26] Over this period there are over 144,000 unique tweets; 26 percent of these tweets are re-tweeted and 33 percent are "liked". Among tweets that are re-tweeted, the average number of re-tweets is 3.9 and the average number of likes is 5.5. For each tweet I observe when the tweet was posted, who posted the tweet, the entire text of the tweet, and the number of likes and re-tweets a tweet has. Unfortunately I am not able to access geographic information about where the twitter user was when they posted the tweet in the historic data and I cannot conduct the same analysis as before. However, I can examine the frequency at which tweets are posted over time and which words are most often associated with the EITC. As seen in Appendix Figure A7, Tweets about the EITC peak during the fourth and fifth week of the year (last week of January, first week of February). This corresponds to the time when most EITC returns are filed and received (Hoynes et al., 2015). Re-tweets and likes of EITC-related

[26]In England the Everton Football club has the handle @eitc (Everton in the Community). I exclude any tweets from @eitc or that mention @eitc. This is about ten percent of the sample. The pictures are similar if I include these tweets.

tweets also spike during this period. As evidence that information can spread to many people through Twitter, there are other spikes in Appendix Figure A7 of likes and re-tweets that are associated with single tweet events.

In Figure 6 I take every word that occurs in the text of these tweets and plot it by the number of tweets it occurs in and the number of times the tweet containing that word is re-tweeted or liked. I exclude the 100 most common English words (e.g., "a", "about", "and", etc) as well as other common words such as "is" and "are" (Appendix Figure A8 shows the same re-tweet plot with successively lower number of tweet thresholds so that less frequently tweeted words can be seen). From Figure 6 several patterns arise shedding light on the types of information about the EITC that are conveyed through tweets. The word that is mentioned the most is "eitc" followed by "#eitc", "tax", "credit", "income", "working", and "families". These tweets are also frequently re-tweeted and liked. In addition to the common words, there are other things common in many of the tweets. Approximately half of the tweets include a website link to EITC guides, assistance groups, and Facebook groups where a viewer can access more information. These tweets generate nearly 100,000 re-tweets and 145,000 likes. Another element common to many of the tweets is a direct mention of another twitter user (using the user's twitter handle) or the inclusion of a picture or graphic. These patterns of transmission are consistent with the twitter social network introducing people to the EITC and providing information, but also directing them to other information sources.

I next turn to Google Trends to look at how people's search interest in EITC related terms varies across time and geography. Google Trends reports the search frequency of a specified term over time. Google Trend data is not available at the county-level, but is available for designated market areas (DMA). A DMA is a group of counties that are meant to capture a media market. A DMA can cross state borders, and I am not able to separate by both DMA and state. Within a given DMA-level query, the month with the highest number of searches for a specified term (e.g., "eitc") is assigned a value of 100. Each other month is then assigned a value between 0 and 100, proportional to the number of searches in that DMA for "eitc" relative to the maximum month. This measure is defined for the given DMA and period of time, so the measures are not directly comparable when making a query for a different geographic area.[27]

To exploit this data, I will adopt a slightly different strategy. First, I pull the entire monthly series for each DMA for several terms related to the EITC ("eitc", "eic", "earned income tax credit", and "earned income credit") as well as terms related to self employment ("self employed" and "schedule c"). I include "eic" and "earned income credit" because this is the terminology used by the IRS, and likely to be the terms used by those first encountering the IRS EITC form (publication 596). As seen in Figure A9, when averaging across all DMA in the US, there is sharp cyclicality in search interest for EITC related terms, with interest peaking in January or February, with almost no searches in the post-tax season months. Search interest for "self employment" also peaks during tax season each year, but is more stable throughout the year. Unlike "self employment", search interest for "schedule c" is much more concentrated among months during the tax season.

To see how social network exposure to state EITCs affect this search behavior, I aggregate friendship links from the county-to-county to the DMA-to-DMA level and explore the impact of network exposure on Google search popularity as follows

$$Y_{dmt} = \sum_{\tau=1}^{11} \beta_\tau \big((\textit{Network Exposure}_{ct} * 1(month = \tau) \big) + \alpha \textit{Network Exposure}_{ct} + \phi_d + \psi_m + \delta_t + \varepsilon_{dmt} \tag{8}$$

where d indexes the DMA, m indexes the month, and t indexes the year. The β_τ coefficients trace out the impact of network exposure for each of the separate months, treating December as the omited month, thus everything is relative to search popularity in December. I also

[27]For example, if one area has perpetually low search volume, a few searches can lead to large differences in this scaled measure. This would work against finding an effect, if search behavior responds in places with more exposure.

include DMA fixed effects, month, fixed effects, and year fixed effects.

In Figure 7 I present the coefficients from these regressions for each month with confidence intervals. For all of the EITC related terms and the self employment related terms, the number of out-of-state friends exposed to a state EITC is associated with an increase in search popularity during January and February. For the term "self employed" and "schedule c" this elevated search popularity continues through the end of tax season in April. Reassuringly, the effects throughout the rest of the year are small and often not statistically different than zero. When looking at the term "eitc" and "earned income tax credit" search interest during non-tax season is statistically different from zero, but actually goes in the opposite direction, consistent with elevated search interest at the end of the year in December. Overall this would suggest that after controlling for DMA level characteristics and aggregate time trends, places with more out-of-state friends exposed to state EITCs see a larger increase in google search interest for EITC and self employment related terms, like "schedule c".[28]

6 Conclusion

There is a sustained interest in understanding how social networks affect people's interactions with government programs and social policies. Until recently, we have had little evidence on how distant social networks affect these decisions. Using the Social Connectedness Index (Bailey et al., 2018b), I explore what happens to EITC filing behavior when households' distant social networks are exposed to a newly implemented state EITC.

Although this policy change does not directly affect the households, I find that the social network's exposure does change EITC filing behavior. When the number of out-of-state friends exposed to a state EITC increases through policy expansions, the EITC filing rate remains constant. However, among EITC filers, there is a shift towards claims that include some self employment income. A one standard deviation increase in the number of out-of-

[28]Google will suppress popularity measures if there is not enough search interest. Unfortunately popularity measures for more specific terms such as "maximize EIC credit" or "bigger EIC credit" are suppressed.

state friends exposed to a state EITC increases the percent of returns claiming the EITC with self employment income by 8.8 percent. Network exposure is also associated with a shifting in the income distribution of households claiming the EITC. Households move away from the low and high ends of the distribution towards the income bins where the EITC credit is the largest. This behavior is consistent with patterns observed by Saez (2010) and Mortenson and Whitten (2018), where households with self employment shift towards refund maximizing points.

These patterns also mimic the behaviors observed among households that were directly affected by the state expansion. I find little evidence that the EITC filing rate changes, but on average when a state introduces an EITC, the EITC filing shifts toward including self employment income by 1.26 percentage points. Also 4.1 percent of the income distribution of EITC claimants shifts toward the middle of the EITC eligibility range. In comparison, social network exposure does not affect the EITC filing rate, and the corresponding effects on self employment filing and the income distribution for a one standard deviation increase in the number of out-of-state friends exposed to a state EITC is about 28 percent as large and the effect transmitted through the social network is only about one fourth as large.

This set of results is consistent with social connectedness and geographic ties conveying information after a policy change, in a way that affects households' interactions with the program. This has implications more broadly about how social networks affect households' interactions with government policy. Changes in policy can generate spillover effects as people learn more about the program through their social networks. Importantly, these spillovers can also make it difficult to evaluate the impact of policies if the social network causes households in the counterfactual group to respond.

References

Aizer, A. and Currie, J. (2004). Networks or neighborhoods? correlations in the use of publicly-funded maternity care in california. *Journal of Public Economics*, 88.

Alstadsaeter, A., Kopczuk, W., and Telle, K. (2018). Social networks and tax avoidance: Evidence from a well-defined norwegian tax shelter. *NBER Working Paper Series*, Working Paper No. 25191.

Autor, D., Dorn, D., and Hanson, G. (2013). The china syndrome: Local labor market effects of import competition in the united states. *American Economic Review*, 103(6):2121–2168.

Bailey, M., Cao, R., Kuchler, T., and Stroebel, J. (2018a). The economic effects of social networks: Evidence from the housing market. *Journal of Political Economy*, 126(6):2224–2276.

Bailey, M., Cao, R., Kuchler, T., Stroebel, J., and Wong, A. (2018b). Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives*, 32(3):259–280.

Bastian, J. and Jones, M. R. (2019). Do eitc expansions pay for themselves? effects on tax revenue and public assistance spending. *Working Paper*.

Bastian, J. and Michelmore, K. (2018). The long-term impact of the earned income tax credit on children's education and employment outcomes. *Journal of Labor Economics*, 36(4):1127–1163.

Baughman, R. A. and Duchovny, N. (2013). State earned income tax credits and the production of child health: Insurance coverage, utilization, and health status. *Working Paper*, University of New Hampshire.

Bertrand, M., Luttmer, E. F., and Mullainathan, S. (2000). Network effects and welfare cultures. *Quarterly Journal of Economics*, 115(3):1019–1055.

Borjas, G. J. and Hilton, L. (1996). Immigration and the welfare state: Immigrant participation in means-tested entitlement programs. *Quarterly Journal of Economics*, 111(2):575–604.

Chetty, R., Friedman, J. N., and Saez, E. (2013). Using differences in knowledge across neighborhoods to uncover the impacts of the eitc on earnings. *The American Economic Review*, 103(7):2683–2721.

Chetty, R. and Saez, E. (2013). Teaching the tax code: Earnings responses to an experiment with eitc recipients. *American Economic Journal: Applied Economics*, 5(1):1–31.

Dahl, G. B., Kostol, A. R., and Mogstad, M. (2014a). Family welfare cultures. *Quarterly Journal of Economics*.

Dahl, G. B., Loken, K. V., and Mogstad, M. (2014b). Peer effects in program participation. *American Economic Review*, 104(7):2049–2074.

Duflo, E. and Saez, E. (2003). The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. *Quarterly Journal of Economics*, 118(3):815–842.

Gee, L. K., Jones, J., and Burke, M. (2017). Social networks and labor markets: How strong ties relate to job finding on facebook's social network. *Journal of Labor Economics*, 35(2):485–518.

Hellerstein, J. K., McInerney, M., and Neumark, D. (2011). Neighbors and co-workers: The importance of residential labor market networks. *Journal of Labor Economics*, 29(4):659–695.

Hoynes, H., Miller, D., and Simon, D. (2015). Income, the earned income tax credit, and infant health. *American Economic Journal: Economic Policy*, 7(1):172–211.

IRS (2018). 2018 tax statistics. *IRS Statistics of Income*.

Kuka, E. (2014). Eitc and the self-employed: Real or reporting effects? *Public Finance Review*, (6):691–719.

Lalumia, S. (2009). The earned income tax credit and reported self-employment income. *National Tax Journal*, 92(7):1698–1719.

Leigh, A. (2010). Who benefits from the earned income tax credit? incidence among recipients, coworkers and firms. *The B.E. Journal of Economic Analysis and Policy*, 10(1).

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60(3):531–542.

Markussen, S. and Roed, K. (2015). Social insurance networks. *Journal of Human Resources*, 50(4):1081–1113.

Michelmore, K. and Lim, K. (2018). The eitc and self-employment among married mothers. *Working Paper*.

Mortenson, J. and Whitten, A. (2018). Bunching to maximize tax credits: Evidence from kinks in the u.s. tax schedule. *Available at SSRN: https://ssrn.com/abstract=2719859*.

Neumark, D. and Wascher, W. (2001). Using the eitc to help poor families: New evidence and a comparison with the minimum wage. *National Tax Journal*, 64(4):712–746.

Neumark, D. and Williams, K. E. (2019). Do state earned income tax credits increase participation in the federal eitc? *Working Paper*.

Nichols, A. and Rothstein, J. (2016). The earned income tax credit. *In Economics of Means-Tested Transfer Programs in the United States*, Volume I(Robert A., Moffit, ed. Chicago: University of Chicago Press).

Rubin, D. (1986). "which ifs have causal answers?" discussion of holland's "statistics and causal inference". *Journal of the American Statistical Association*, 81.

Saez, E. (2010). Do taxpayers bunch at kink points. *American Economic Journal: Economic Policy*, 2.

Wilson, R. (2018). The eitc and employment transitions: Labor force attachment and annual exit. *Working Paper*.

Wolfers, J. (2006). Did unilateral divorce laws raise divorce rates? a reconciliation and new results. *American Economic Review*, 96(5):1802–1820.

Tables and Figures

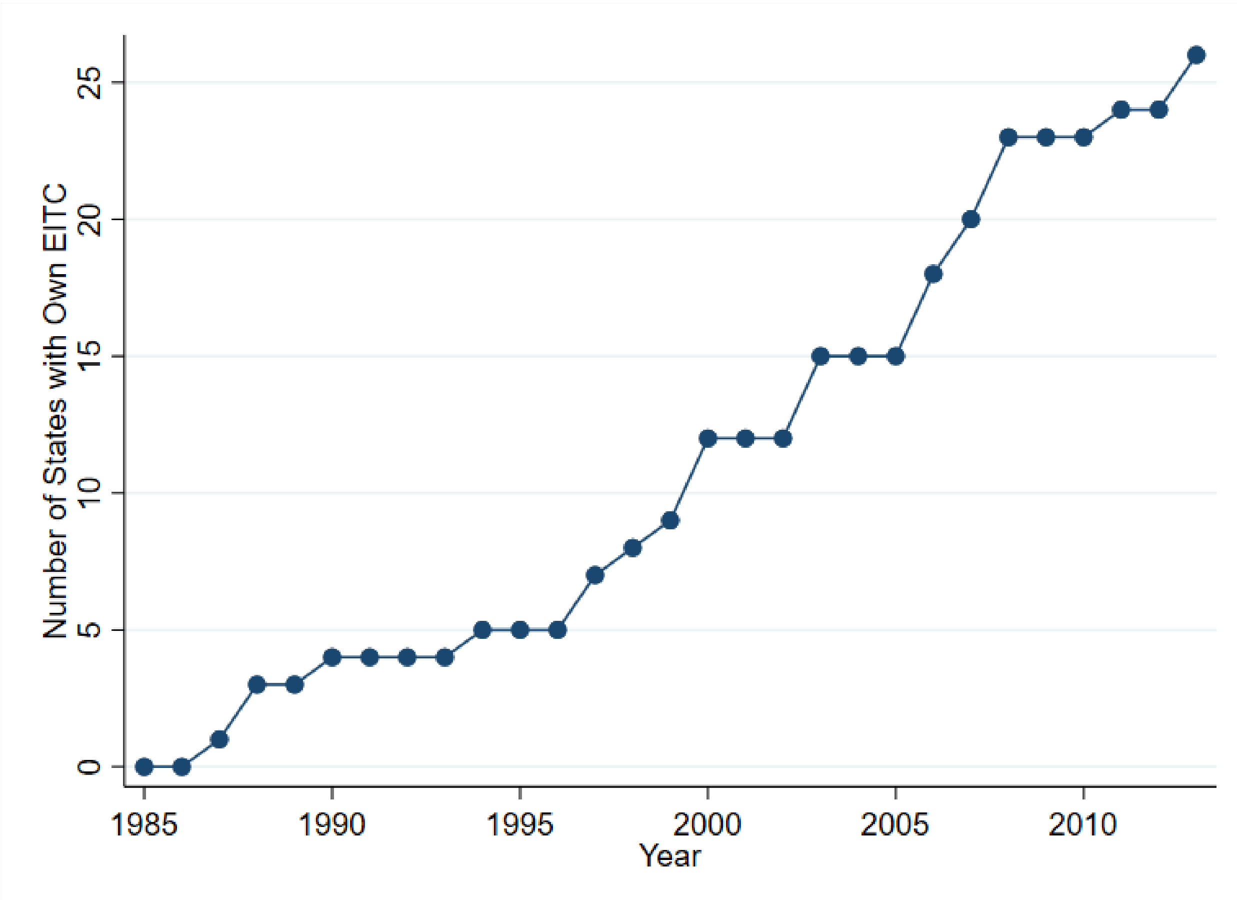

Figure 1: Fraction of States with A State EITC

Source: Author's own calculations.

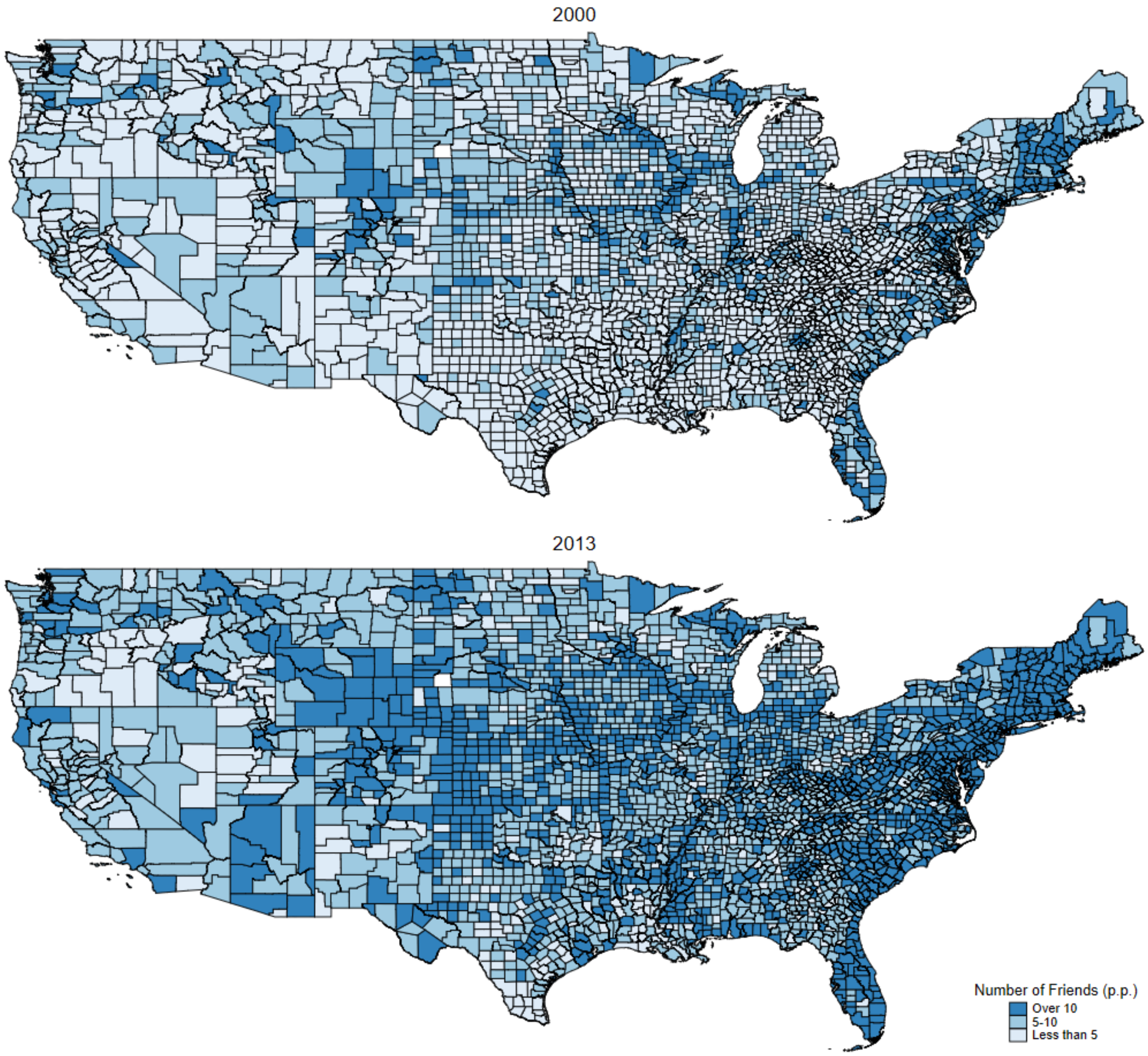

Figure 2: Number of Out-of-State Friends per Person Exposed to State EITC in 2000 and 2013

Source: Author's own calculations. Out-of-state friend links obtained from the Social Connectedness Index, an dataset of anonymized Facebook friendship links reported at the county pair level (Bailey et al., 2018b). This is a scalar multiple of the actual number of Facebook friends, which is not reported. Exposure to state EITC is constructed by linking the Social Connectedness Index to state-level roll out of EITC programs.

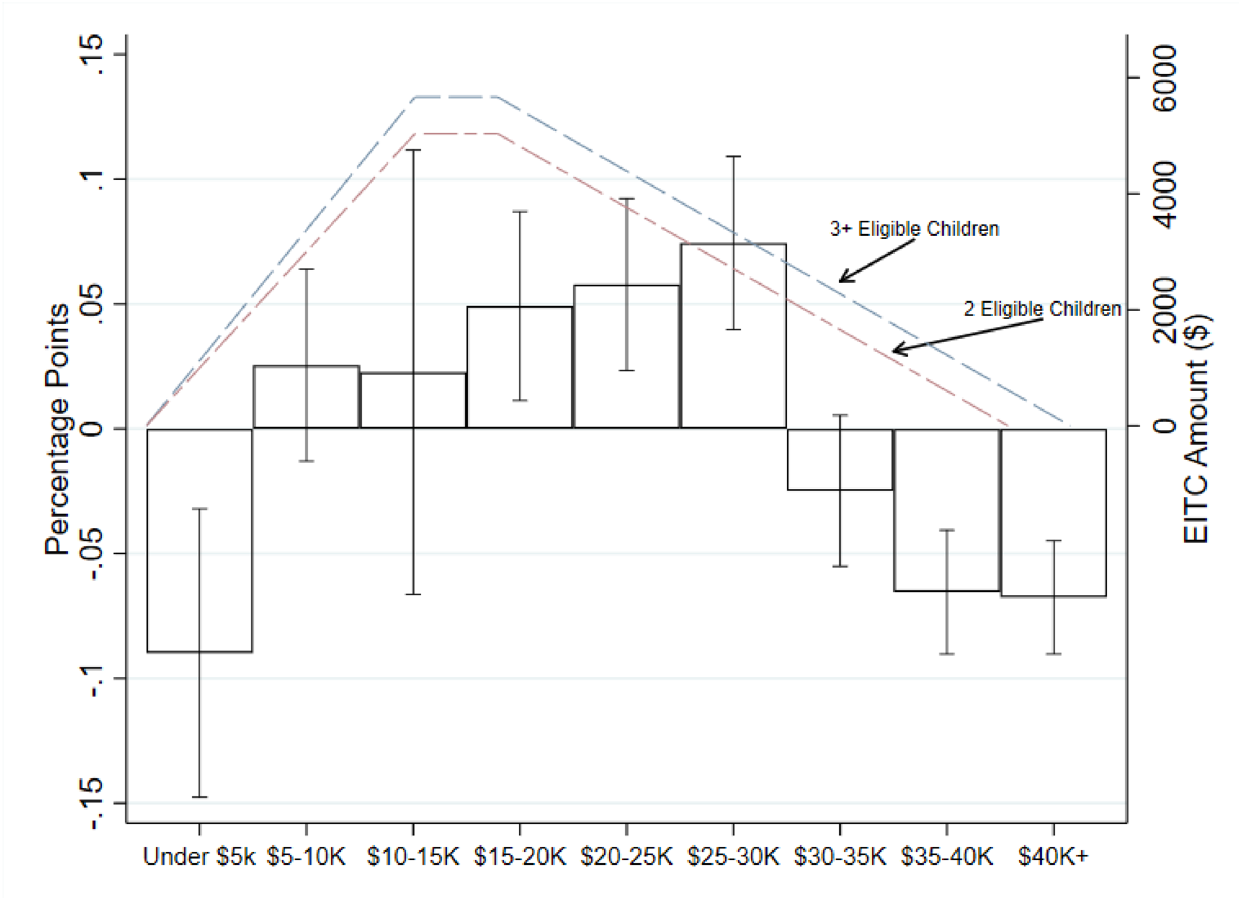

Figure 3: Marginal Impact of State EITC on Earnings Distribution of EITC Recipients for a One Percent Increase in the 1999 EITC Filing Rate

Notes: Each bar represents the coefficient from regression equation (2), and is the marginal impact of having a state EITC for a one percent increase in the 1999 EITC filing rate. For reference, the average 1999 EITC filing rate was 18 percent. The outcomes are the percent of EITC filing units in each $5,000 bin in the income distribution through 35 thousand dollars, and for 40 thousand dollars or more. Standard errors are corrected for clustering at the state level. Ninety five percent confidence intervals are provided. The EITC parameters for a family with two or three eligible children is also provided for reference.

Source: Author's own calculations. EITC claiming data obtained from the Brookings Institution IRS EITC interactive dataset. EITC parameters obtained from the Tax Policy Center.

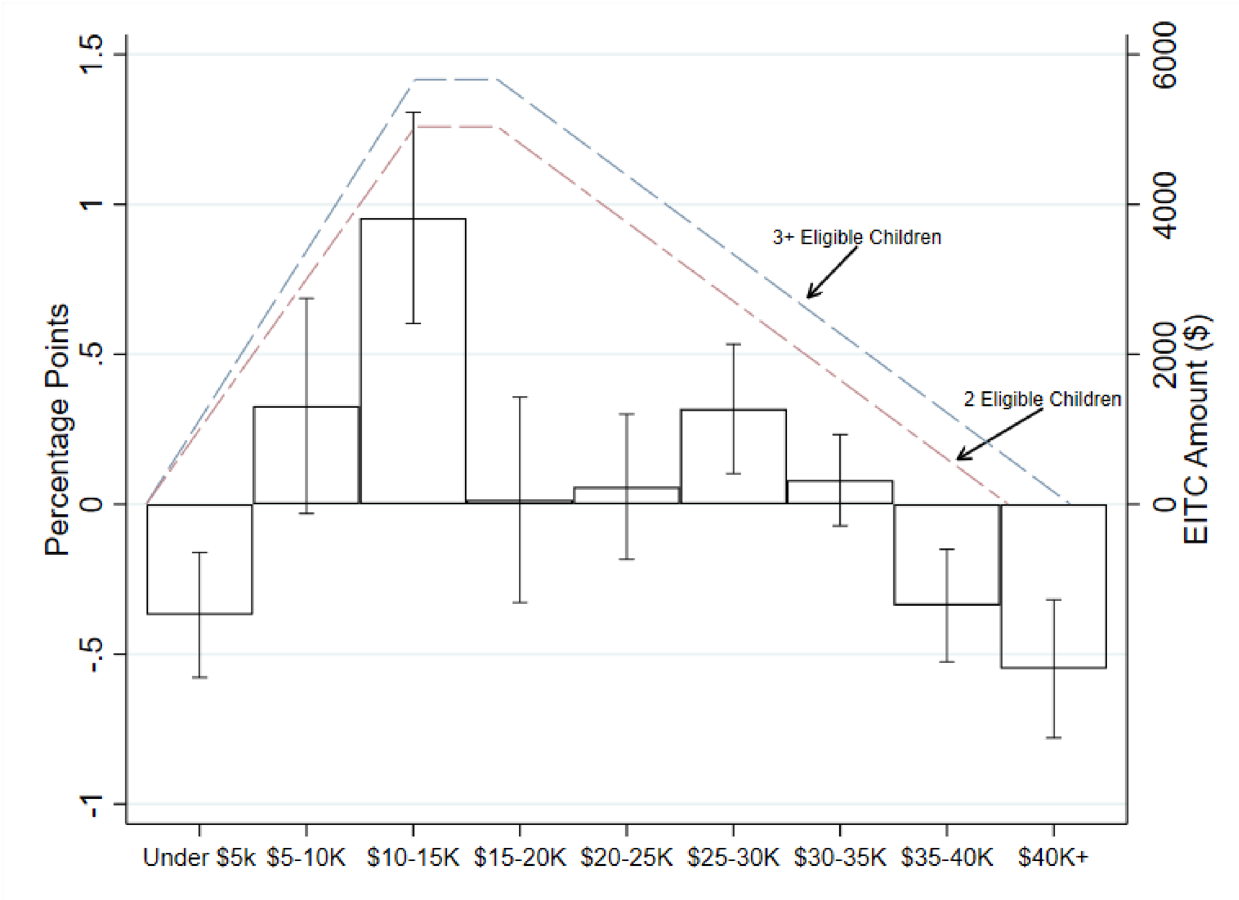

Figure 4: Marginal Impact of a One Standard Deviation Increase in the Number of Out-of-State Friends per Person Exposed to a State EITC on Earnings Distribution of EITC Recipients

Notes: Each bar represents the coefficient from regression equation (4), and represent the impact of a one standard deviation increase in the number of out-of-state friends per person exposed to a state EITC. The outcomes are the percent of EITC filing units in each $5,000 bin in the income distribution through 35 thousand dollars, and for 40 thousand dollars or more. Standard errors are corrected for clustering at the state level. Ninety five percent confidence intervals are provided. The EITC parameters for a family with two or three eligible children is also provided for reference.

Source: Author's own calculations. Out-of-state friend links obtained from the Social Connectedness Index, an dataset of anonymized Facebook friendship links reported at the county pair level (Bailey et al., 2018b). EITC claiming data obtained from the Brookings Institution IRS EITC interactive dataset. EITC parameters obtained from the Tax Policy Center.

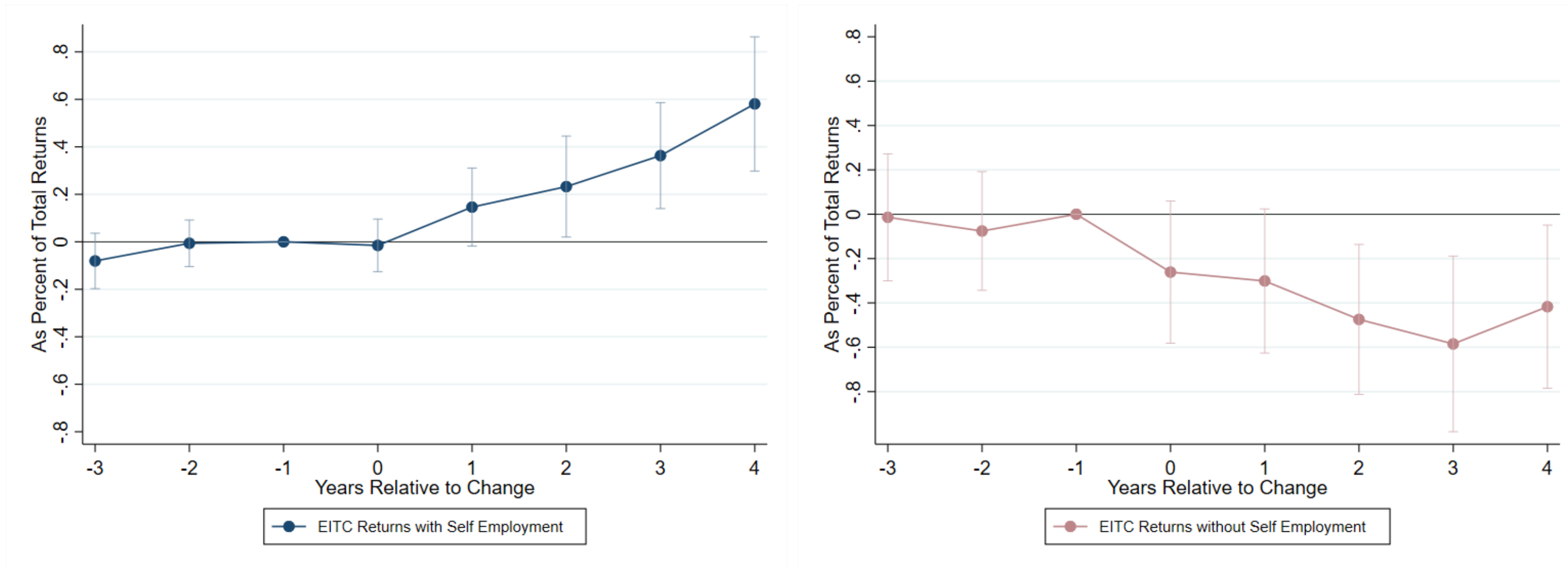

Figure 5: State Implementation Event Study: Impact of the Number of Out-of-State Friends per Person Exposed to New State EITC on the Percent of Returns with EITC

Notes: Each plot represents the coefficient from regression equation (6), where the outcomes are the percent of returns claiming the EITC with self employment (Schedule C, E, or F) and the percent of returns claiming the EITC without self employment (no Schedule C, E, or F). Standard errors are corrected for clustering at the state level. Standard errors are similar if correcting for two-way clustering at the state level and the event state level. Ninety five percent confidence intervals are provided.

Source: Author's own calculations. Out-of-state friend links obtained from the Social Connectedness Index, a dataset of anonymized Facebook friendship links reported at the county pair level (Bailey et al., 2018b). EITC claiming data obtained from the Brookings Institution IRS EITC interactive dataset.

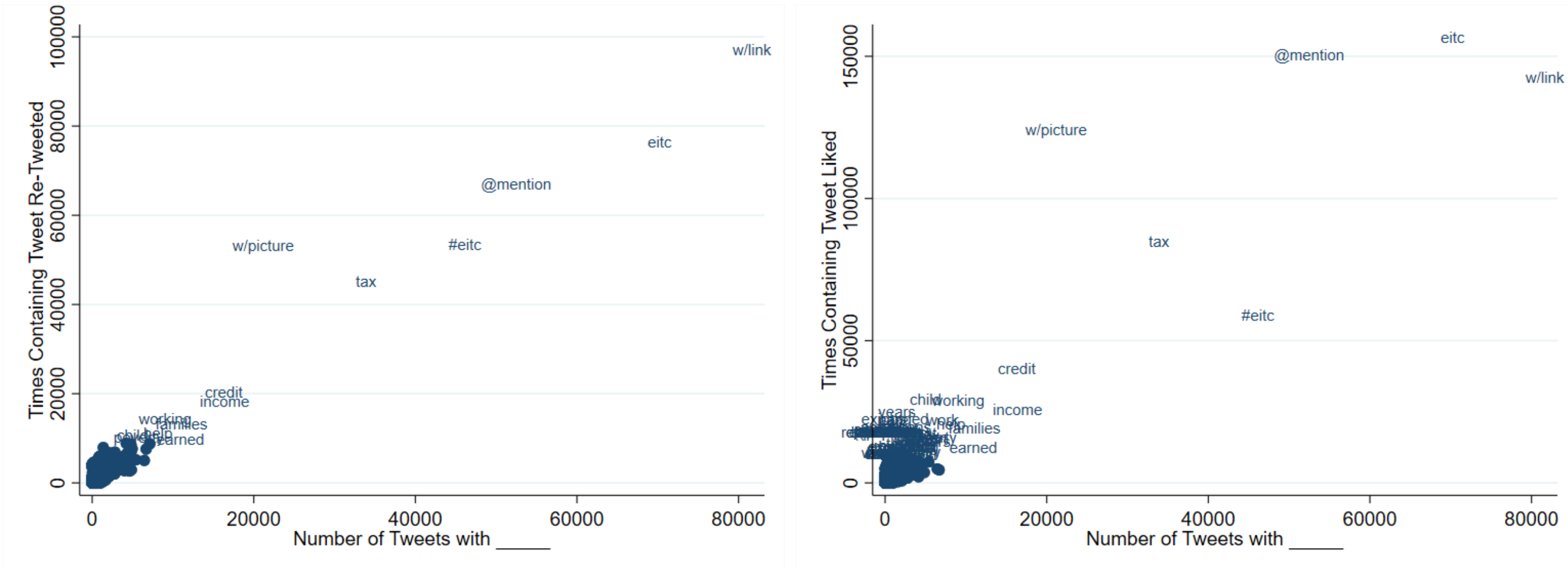

Figure 6: Re-Tweet and Like Frequency by Word Count Frequency, January 2010-September 2019

Notes: Tweets since 2010 containing "EITC" were scraped from Twitter in September 2019. Tweets referencing the English Football club with handle @EITC were excluded. The tweet count in the first panel does not include re-tweets. The spike in likes and re-tweets in week 25 are due to three tweets from Noah Smith about the EITC. The spikes in week 46 are due to a single tweet by Representative Joe Kennedy. The dashed line indicates the level when these four tweets are excluded. The vertical black dotted line marks the week of April 15, when federal taxes are due.

Source: Author's own calculations. Tweets posted between January 2010 and September 2019 taken from Twitter.

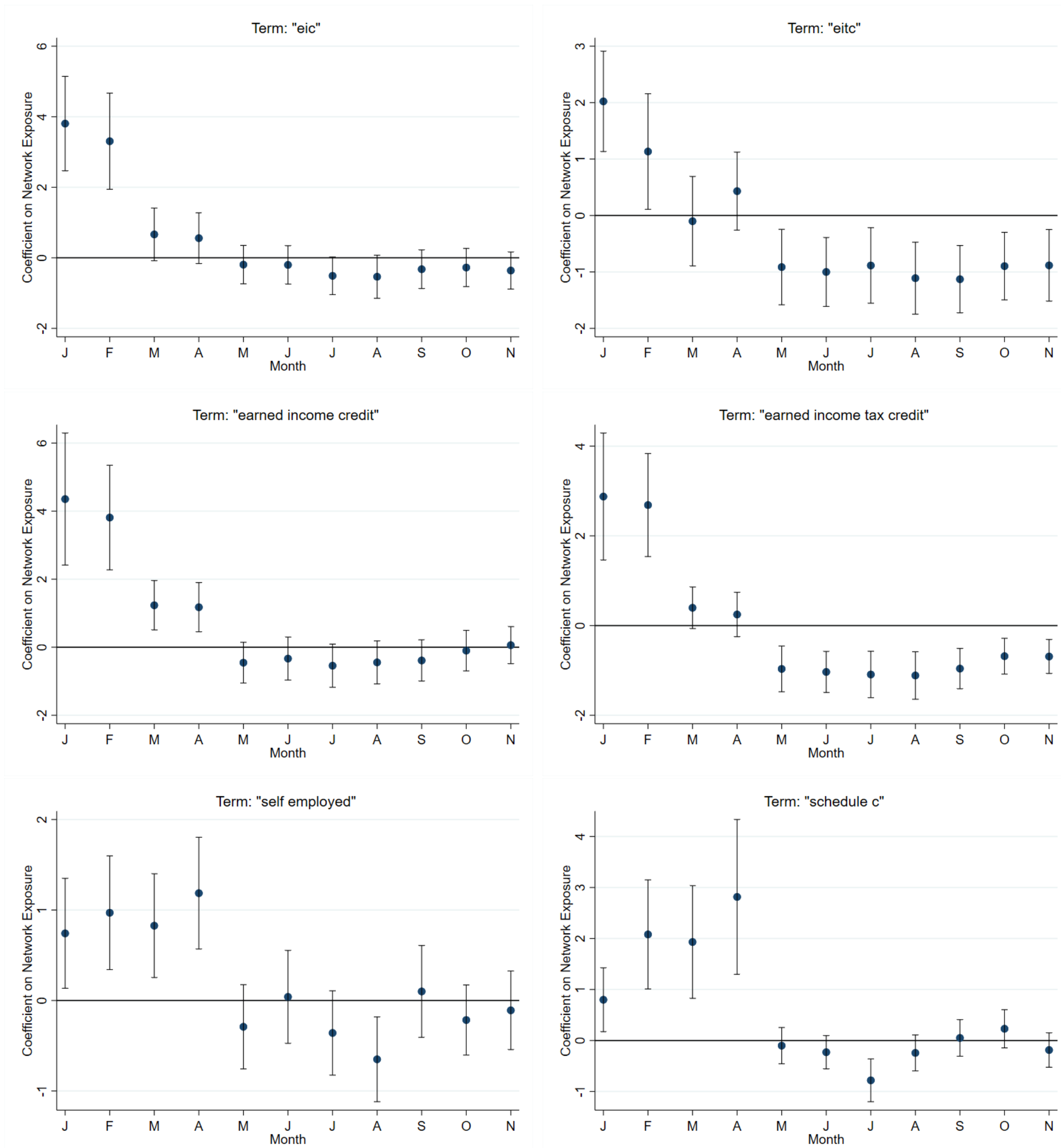

Figure 7: Google Search Intensity for EITC terms by the Number of Out-of-State Friends Exposed to State EITCs

Notes: Each figure represents the coefficients from the regression in equation (8) where the outcome is the Google search index for the listed term and the variable of interest is the standardized number of out-of-state friends exposed to state EITCs interacted with month dummies. DMA, month, and year fixed effects are also included. For each DMA, the month with the highest search intensity is given a value of 100. Every other month is given a score between 0 and 100, proportional to the number of searches relative to the maximum. DMAs cross state lines. 95 Percent Confidence Intervals provided using standard errors corrected for clustering at the DMA level.

Source: Author's own calculations. Out-of-state friend links obtained from the Social Connectedness Index, a dataset of anonymized Facebook friendship links reported at the county pair level (Bailey et al., 2018b). Google search intensity taken from Google Trends at the designated market area level by month level.

Table 1: County-Level Summary Statistics

	Mean in 2000		Δ from 2000 to 2013	
	Change in Number of Friends Exposed		Change in Number of Friends Exposed	
	Below State Median	Above State Median	Below State Median	Above State Median
	(1)	(2)	(3)	(4)
Out-of-State Friends Exposed to State EITC in 2000	6.85	9.76	4.01	6.68
Percent of Returns with EITC	16.72	14.95	4.82	5.30
No-Self Employment	13.36	12.43	2.70	2.79
Self Employment	3.36	2.52	2.11	2.50
Average EITC Amount	1615.45	1614.66	708.36	734.63
Percent Female	0.51	0.52	-0.00	-0.00
Percent NH White	0.76	0.71	-0.05	-0.06
Percent NH Black	0.08	0.13	0.01	0.01
Percent Hispanic	0.12	0.11	0.03	0.03
Percent NH Other	0.04	0.05	0.01	0.02
Percent Less HS	22.70	18.68	-6.65	-5.20
Percent HS	32.20	27.44	0.17	-0.42
Percent Some College	25.41	28.02	2.72	1.18
Percent College	19.69	25.85	3.76	4.44
Unemployment Rate	4.51	3.92	3.47	3.47
Average Job Earnings	43267.22	49106.86	1426.18	2376.37
Observations	1,532	1,556	1,532	1,556

Notes: Columns (1) and (2) report the average county level measures in 2000 for counties where the change in the number of friends exposed to a state EITC was below (Column 1) and above (Column 2) the state median. Columns (3) and (4) report the average change in the county level measures from 2000 to 2013 for counties where the change in the number of friends exposed to a state EITC was below (Column 3) and above (Column 4) the state median.

Table 2: Impact of State EITC on EITC Filing Behavior

	Percent of Returns with EITC			
	Any EITC	No Self Employment (No Schedule C, E, or F)	Self Employment (Schedule C, E, or F)	Percent of EITC Returns with Self Emp. (Schedule C, E, or F)
	(1)	(2)	(3)	(4)
Any State EITC*Share HH in 1999 Claiming EITC	0.01	-0.06***	0.07**	0.21**
	(0.03)	(0.02)	(0.03)	(0.09)
Dependent Mean	18.1	14.0	4.1	22.9
Mean 1999 EITC Rate	15.4	15.4	15.4	15.4
Observations	43,232	43,232	43,232	43,232

Notes: Observation at the county by year level from 2000 to 2013. EITC claiming data obtained from the Brookings/IRS EITC interactive. Estimation controls for the county-level gender and race shares. Observations are weighted by the 2000 county population. County fixed effects are included to control for time invariant county characteristics. State-by-year fixed effects are also included to make this a comparison between counties in the same state. Standard errors are corrected for clustering at the state level. <0.01 ***, p<0.05**, p<0.1*.

Table 3: Impact of State EITC on Income Distribution of EITC Filers

		Percent of EITC Returns by Income Bin								
	Average Refund	Under $5K	$5-10K	$10-15K	$15-20K	$20-25K	$25-30K	$30-35K	$35-40K	Over $40K
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Any State EITC*Share HH	6.13***	-0.09***	0.03	0.02	0.05**	0.06***	0.07***	-0.02	-0.07***	-0.07***
in 1999 Claiming EITC	(1.52)	(0.03)	(0.02)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
Dependent Mean	1953.9	12.2	20.3	18.9	14.2	12.6	10.3	6.0	2.4	1.2
Mean 1999 EITC Rate	15.4	15.4	15.4	15.4	15.4	15.4	15.4	15.4	15.4	15.4
Observations	43,232	43,232	43,232	43,232	43,232	43,232	43,232	43,232	43,232	43,232

Notes: Observation at the county by year level from 2000 to 2013. EITC claiming data obtained from the Brookings/IRS EITC interactive. Estimation controls for the county-level gender and race shares. Observations are weighted by the 2000 county population. County fixed effects are included to control for time invariant county characteristics. State-by-year fixed effects are also included to make this a comparison between counties in the same state. Standard errors are corrected for clustering at the state level. <0.01 ***, p<0.05**, p<0.1*.

Table 4: Impact of Out-of-State Friend Exposure to State EITC on EITC Filing Behavior

	Percent of Returns with EITC			
	Any EITC	No Self Employment (No Schedule C, E, or F)	Self Employment (Schedule C, E, or F)	Percent of EITC Returns with Self Emp. (Schedule C, E, or F)
	(1)	(2)	(3)	(4)
Out-of-State Friends per person	0.02	-0.35**	0.36**	2.32***
Exposed to State EITC (Standardized)	(0.14)	(0.15)	(0.14)	(0.50)
Dependent Mean	18.1	14.0	4.1	22.9
Observations	43,349	43,349	43,349	43,349

Notes: Observation at the county by year level from 2000 to 2013. Out-of-State Friendship data constructed from the Social Connectedness Index (Bailey et al., 2018b). EITC claiming data obtained from the Brookings/IRS EITC interactive. Estimation controls for the county-level gender and race shares. Observations are weighted by the 2000 county population. County fixed effects are included to control for time invariant county characteristics. State by year fixed effects are included, making this a within state comparison. Standard errors are corrected for clustering at the state level. <0.01 ***, p<0.05**, p<0.1*.

Table 5: Impact of Out-of-State Friend Exposure to State EITC on Income Distribution of EITC Filers

		Percent of EITC Returns by Income Bin								
	Average Refund	Under $5K	$5-10K	$10-15K	$15-20K	$20-25K	$25-30K	$30-35K	$35-40K	Over $40K
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Out-of-State Friends per person	0.60	-0.37***	0.33*	0.96***	0.01	0.06	0.32***	0.08	-0.34***	-0.55***
Exposed to State EITC (Standardized)	(7.45)	(0.10)	(0.18)	(0.18)	(0.17)	(0.12)	(0.11)	(0.08)	(0.09)	(0.12)
Dependent Mean	1954.0	12.2	20.3	18.9	14.2	12.6	10.3	6.0	2.4	1.2
Observations	43,349	43,349	43,349	43,349	43,349	43,349	43,349	43,349	43,349	43,349

Notes: Observation at the county by year level from 2000 to 2013. Out-of-State Friendship data constructed from the Social Connectedness Index (Bailey et al., 2018b). EITC claiming data obtained from the Brookings/IRS EITC interactive. Estimation controls for the county-level gender and race shares. Observations are weighted by the 2000 county population. County fixed effects are included to control for time invariant county characteristics. State by year fixed effects are included, making this a within state comparison. Standard errors are corrected for clustering at the state level. <0.01 ***, p<0.05**, p<0.1*.

Table 6: Impact of Out-of-State Friend Exposure to State EITC on Family-level Reported Self-Employment (ACS 2005-2017)

		Self-Employment Rate		
	Employment Rate	Any	Not Incorporated	Incorporated
	(1)	(2)	(3)	(4)
Out-of-State Friends per person	0.14	0.20*	0.23**	-0.02
Exposed to State EITC (Standardized)	(0.17)	(0.12)	(0.10)	(0.06)
Dependent Mean	70.70	13.98	10.17	4.08
Observations	11,479	11,479	11,479	11,479

Notes: Observation at the commuting zone/state level from the ACS between 2005 and 2017. Out-of-State Friendship data constructed from the Social Connectedness Index (Bailey et al., 2018b), and aggregated up to the commuting zone/state level. This measure is standardized by the same mean and standard deviation as above, so that coefficients can be directly compared. Estimation controls for area level age group shares, marital status shares, race and ethnicity group shares, education group shares, share of family units with a single mother, and the share of family units with one, two, or more EITC eligible children. Observations are collapsed from the family unit level using the household weights provided in the ACS, where individuals are probabilistically assigned to commuting zones as in (Autor et al., 2013), but the state of residence is maintained resulting in commuting zone by state level geographies. Commuting zone/state geographies are then weighted by 2010 population. Commuting zone/state fixed effects are included to control for time invariant geographic characteristics. State by year fixed effects are included, making this a within state comparison. Standard errors are corrected for clustering at the state level. <0.01 ***, $p<0.05$**, $p<0.1$*.

Table 7: Impact of Out-of-State Friend Exposure to State EITC on EITC Filing Behavior, Predicted Network

	Percent of Returns with EITC			Percent of EITC Returns with Self Emp. (Schedule C, E, or F)
	Any EITC	No Self Employment (No Schedule C, E, or F)	Self Employment (Schedule C, E, or F)	
	(1)	(2)	(3)	(4)
Predicted Out-of-State Friends per person	0.01	-0.22***	0.23***	1.30***
Exposed to State EITC (Standardized)	(0.06)	(0.08)	(0.07)	(0.26)
Dependent Mean	18.1	14.0	4.1	22.9
Observations	43,349	43,349	43,349	43,349

Notes: Observation at the county by year level from 2000 to 2013. Out-of-State Friendship data constructed from the Social Connectedness Index (Bailey et al., 2018b). EITC claiming data obtained from the Brookings/IRS EITC interactive. Estimation controls for the county-level gender and race shares. Observations are weighted by the 2000 county population. County fixed effects are included to control for time invariant county characteristics. State by year fixed effects are included, making this a within state comparison. Standard errors are corrected for clustering at the state level. <0.01 ***, p<0.05**, p<0.1*.

Appendix Tables and Figures

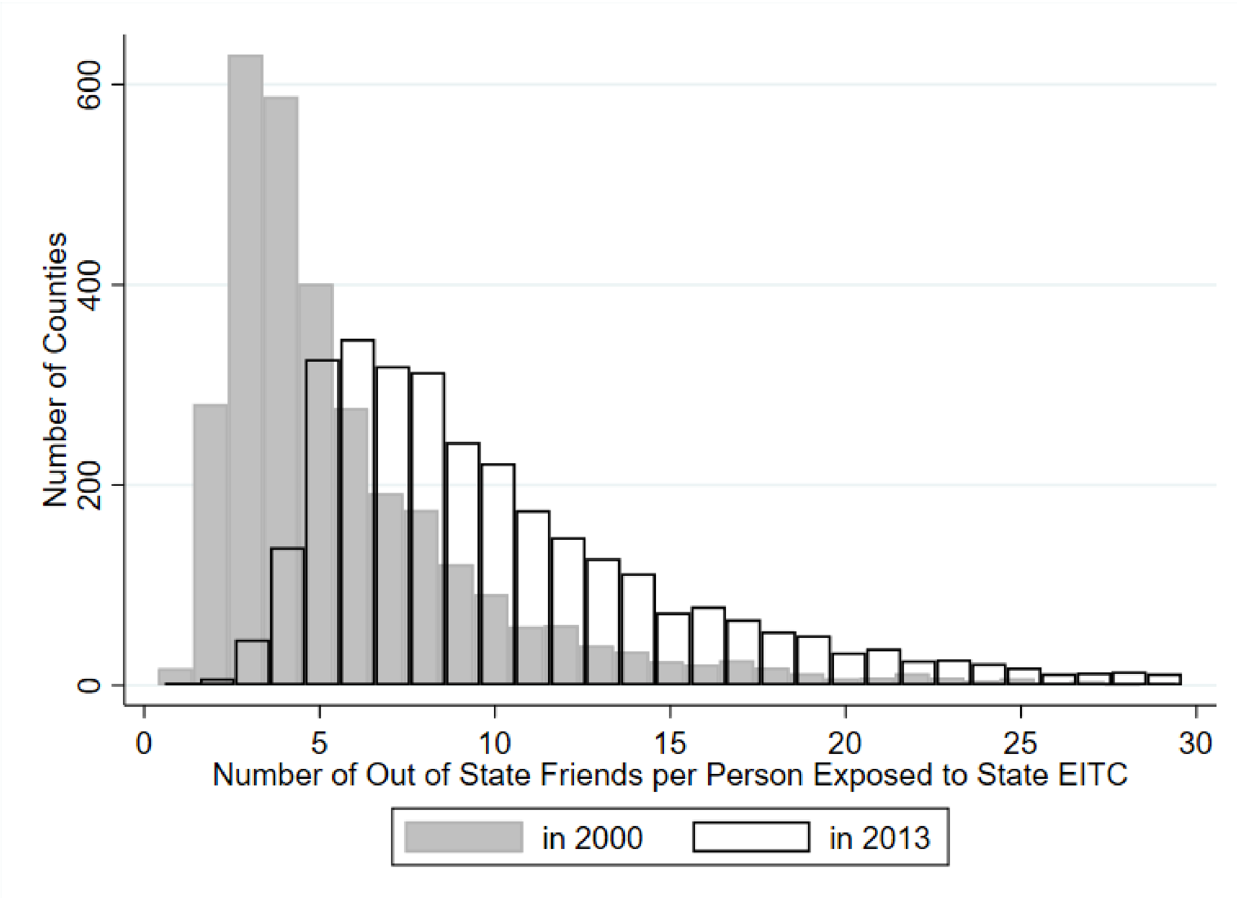

Figure A1: Number of Out-of-State Friends per Person Exposed to State EITC in 2000 and 2013

Source: Author's own calculations. Out-of-state friend links obtained from the Social Connectedness Index, a dataset of anonymized Facebook friendship links reported at the county pair level (Bailey et al., 2018b). This is a scalar multiple of the actual number of Facebook friends, which is not reported. Exposure to state EITC is constructed by linking the Social Connectedness Index to state-level roll out of EITC programs.

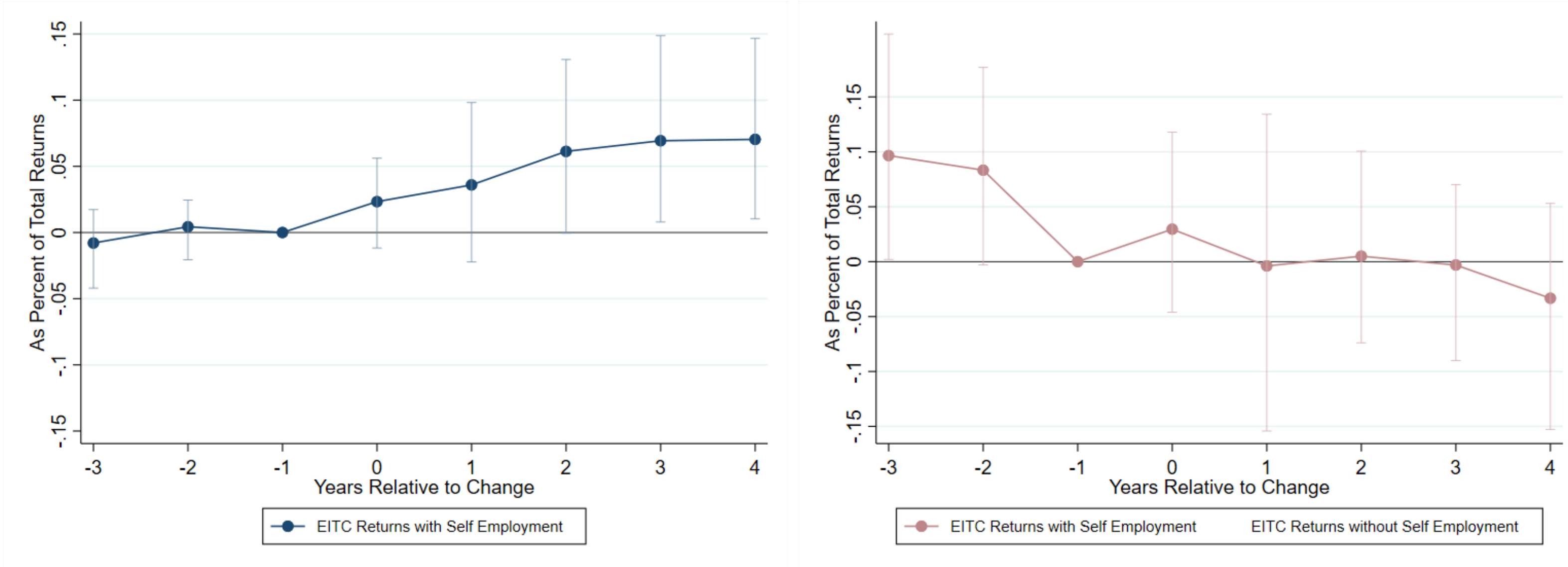

Figure A2: State Implementation Event Study: Direct Impact of the New State EITC on the Percent of Returns with EITC

Notes: Each plot represents the coefficient from regressing the EITC outcome on the county percent of households in the EITC range in 2000 interacted with year indicators which capture how many years from state implementation. County and state by year fixed effects are also included. Only expansion states are included in the regression. The outcomes are the percent of returns claiming the EITC with self employment (Schedule C, E, or F) and the percent of returns claiming the EITC without self employment (no Schedule C, E, or F). Ninety five percent confidence intervals are provided using wild bootstrapped standard errors since there are only 11 state events in the balanced panel between 2003 and 2010.

Source: Author's own calculations. EITC claiming data obtained from the Brookings Institution IRS EITC interactive dataset.

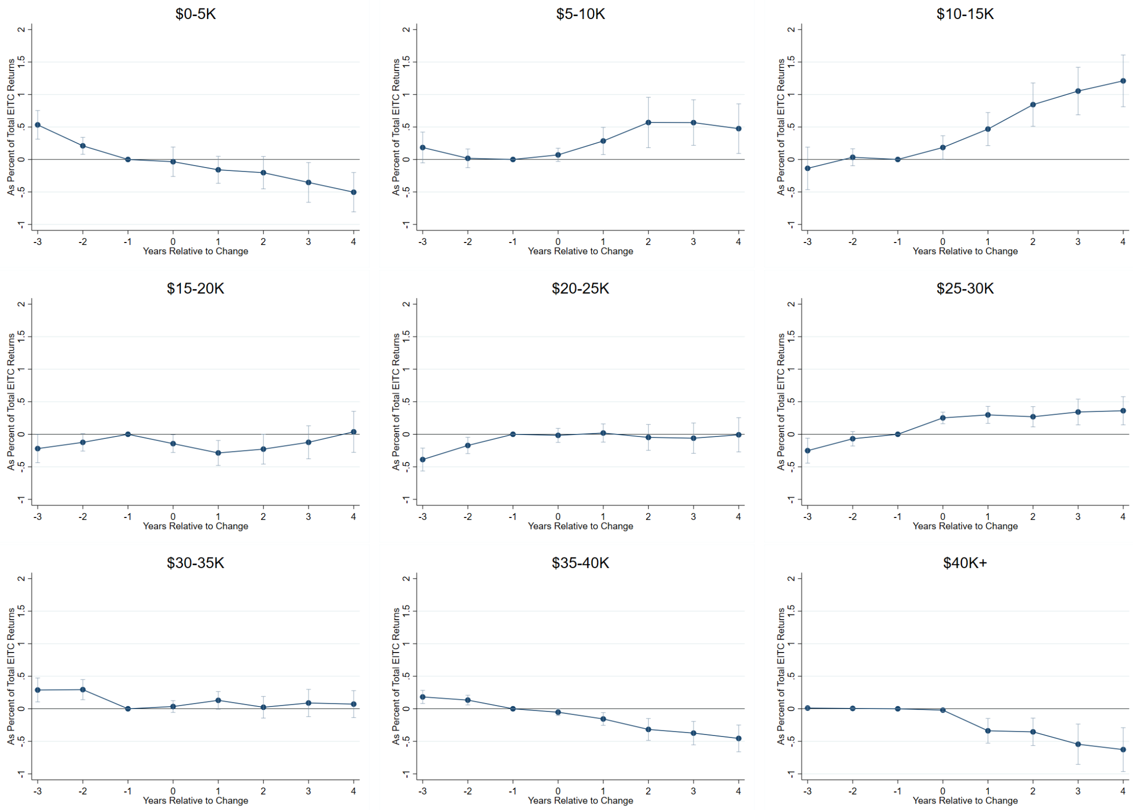

Figure A3: Event Study: Impact of Share of Friends Exposed to State EITC on the Percent of EITC Returns in Each Income Group

Notes: Each plot represents the coefficient from regression equation (6), where the outcomes are the percent of returns claiming the EITC in each income bin. Standard errors are corrected for clustering at the state leve. Ninety five percent confidence intervals are provided.

Source: Author's own calculations. Out-of-state friend links obtained from the Social Connectedness Index, a dataset of anonymized Facebook friendship links reported at the county pair level (Bailey et al., 2018b). EITC claiming data obtained from the Brookings Institution IRS EITC interactive dataset.

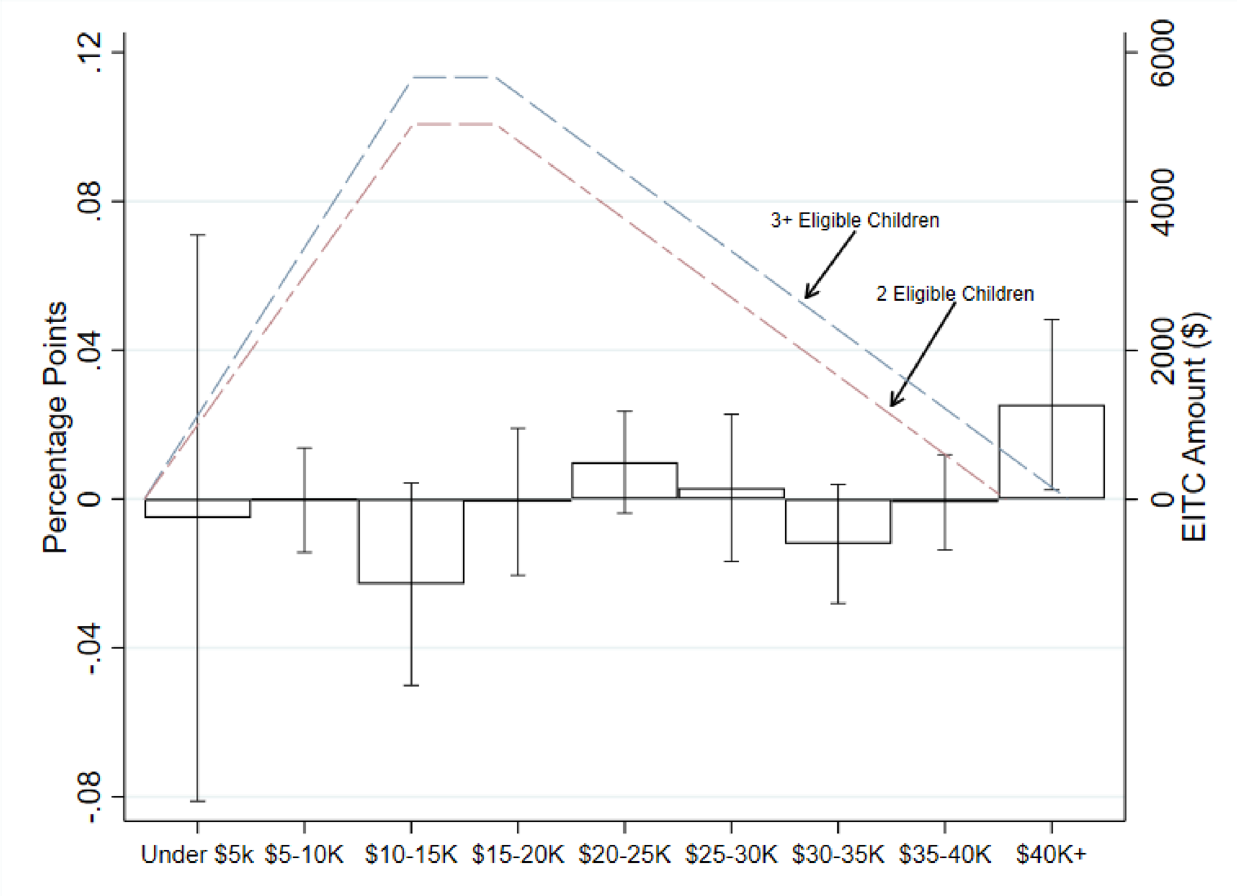

Figure A4: Marginal Impact of State EITC on Earnings Distribution of *Non-EITC* Recipients for a One Percent Increase in the 1999 EITC Filing Rate

Notes: Each bar represents the coefficient from regression equation (4), and is the marginal impact of having a state EITC for a one percent increase in the 1999 EITC filing rate. For reference, the average 1999 EITC filing rate was 18 percent. The outcomes are the percent of non-EITC filing units in each $5,000 bin in the income distribution through 35 thousand dollars, and for 40 thousand dollars or more. Standard errors are corrected for clustering at the state level. Ninety five percent confidence intervals are provided. The EITC parameters for a family with three eligible children is also provided for reference.

Source: Author's own calculations. Out-of-state friend links obtained from the Social Connectedness Index, an dataset of anonymized Facebook friendship links reported at the county pair level (Bailey et al., 2018b). EITC claiming data obtained from the Brookings Institution IRS EITC interactive dataset. EITC parameters obtained from the Tax Policy Center.

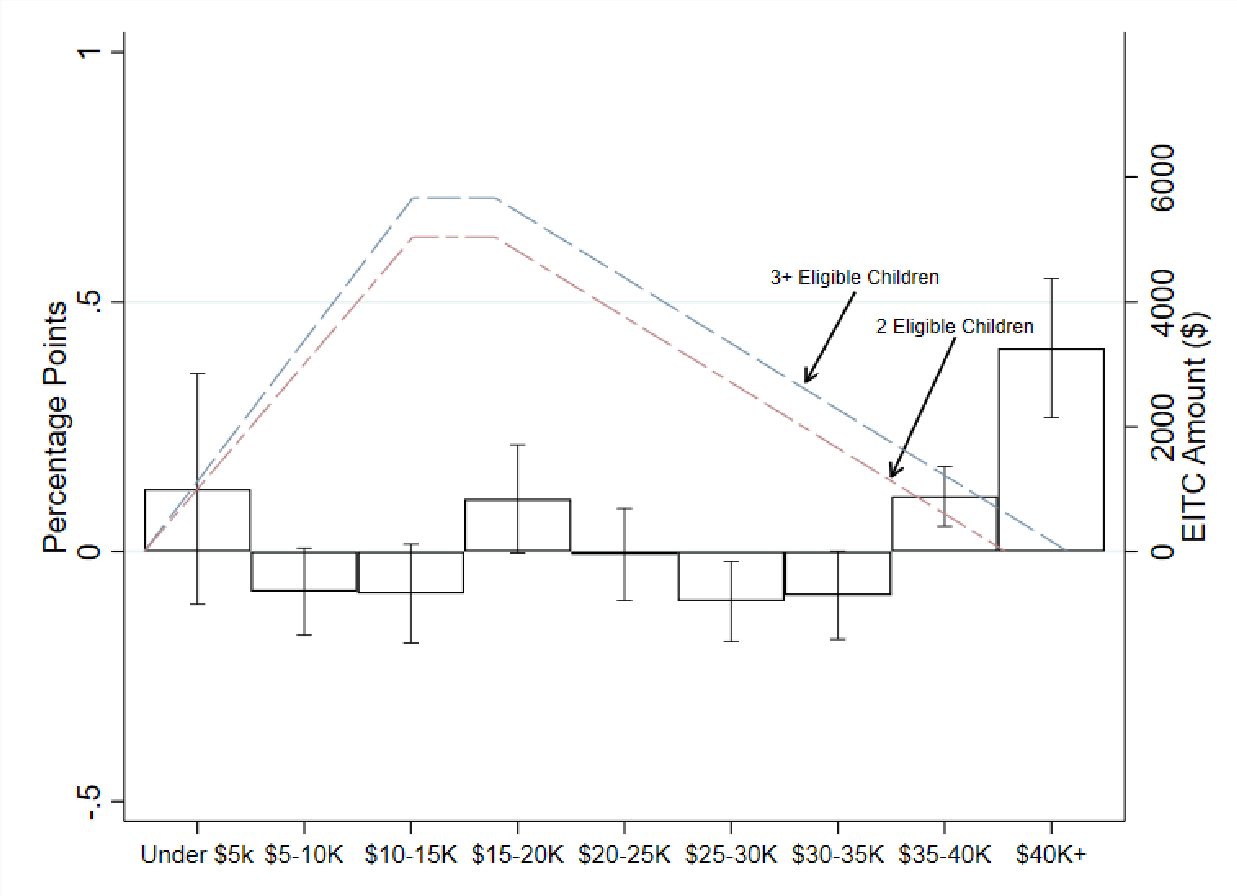

Figure A5: Impact of a One Standard Deviation Increase in the Number of Out-of-State Friends per Person Exposed to State EITC on Earnings Distribution of *Non-EITC* Recipients

Notes: Each bar represents the coefficient from regression equation (4), and is the impact of a one standard deviation increase in the number of out-of-state friends per person exposed to a state EITC. The outcomes are the percent of non-EITC filing units in each $5,000 bin in the income distribution through 35 thousand dollars, and for 40 thousand dollars or more. Standard errors are corrected for clustering at the state level. Ninety five percent confidence intervals are provided. The EITC parameters for a family with three eligible children is also provided for reference.

Source: Author's own calculations. Out-of-state friend links obtained from the Social Connectedness Index, an dataset of anonymized Facebook friendship links reported at the county pair level (Bailey et al., 2018b). EITC claiming data obtained from the Brookings Institution IRS EITC interactive dataset. EITC parameters obtained from the Tax Policy Center.

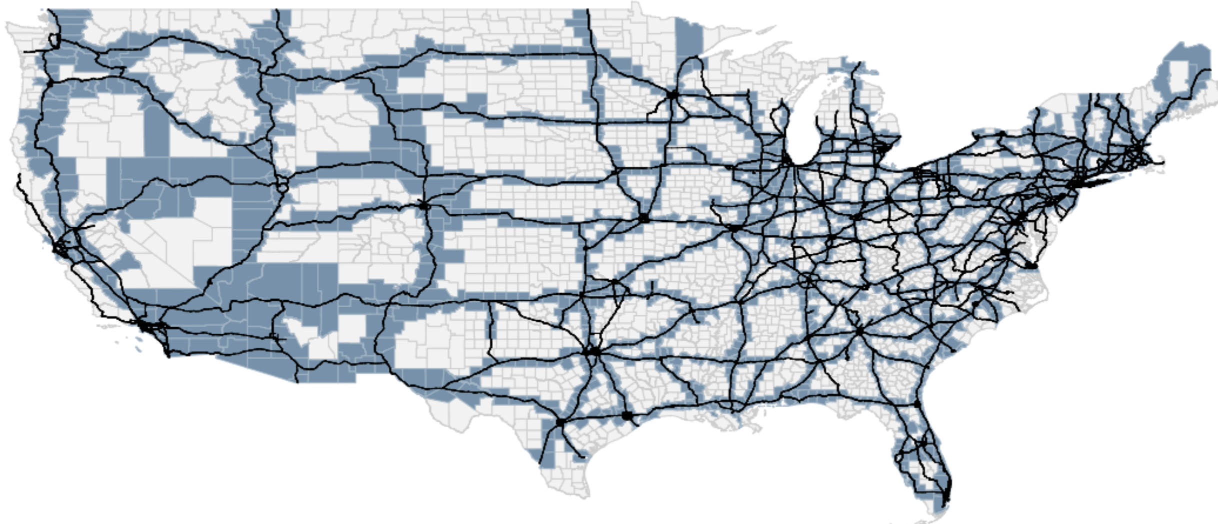

Figure A6: Interstate Routes and Counties with Any Interstate

Source: Author's own calculations. Interstate shape files created using ArcGIS.

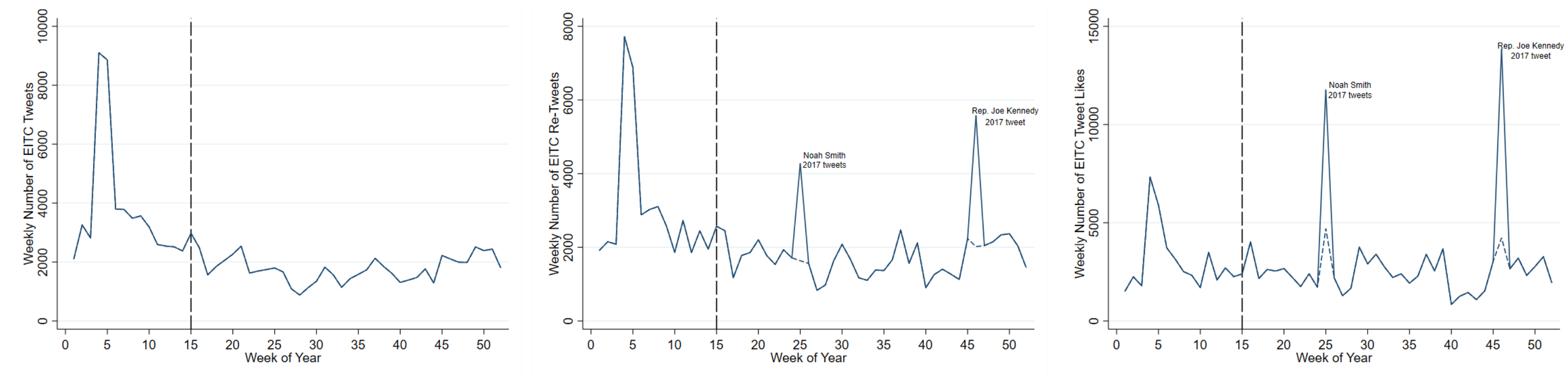

Figure A7: Total EITC Related Tweets, Re-Tweets, and Likes by Week of the Year 2010-2018

Notes: Tweets since 2010 containing "EITC" were scraped from Twitter in September 2019. Tweets referencing the English Football club with handle @EITC were excluded. The tweet count in the first panel does not include re-tweets. The spike in likes and re-tweets in week 25 are due to three tweets from Noah Smith about the EITC. The spikes in week 46 are due to a single tweet by Representative Joe Kennedy. The dashed line indicates the level when these four tweets are excluded. The vertical black dotted line marks the week of April 15, when federal taxes are due.

Source: Author's own calculations. Tweets posted between January 2010 and September 2019 taken from Twitter.

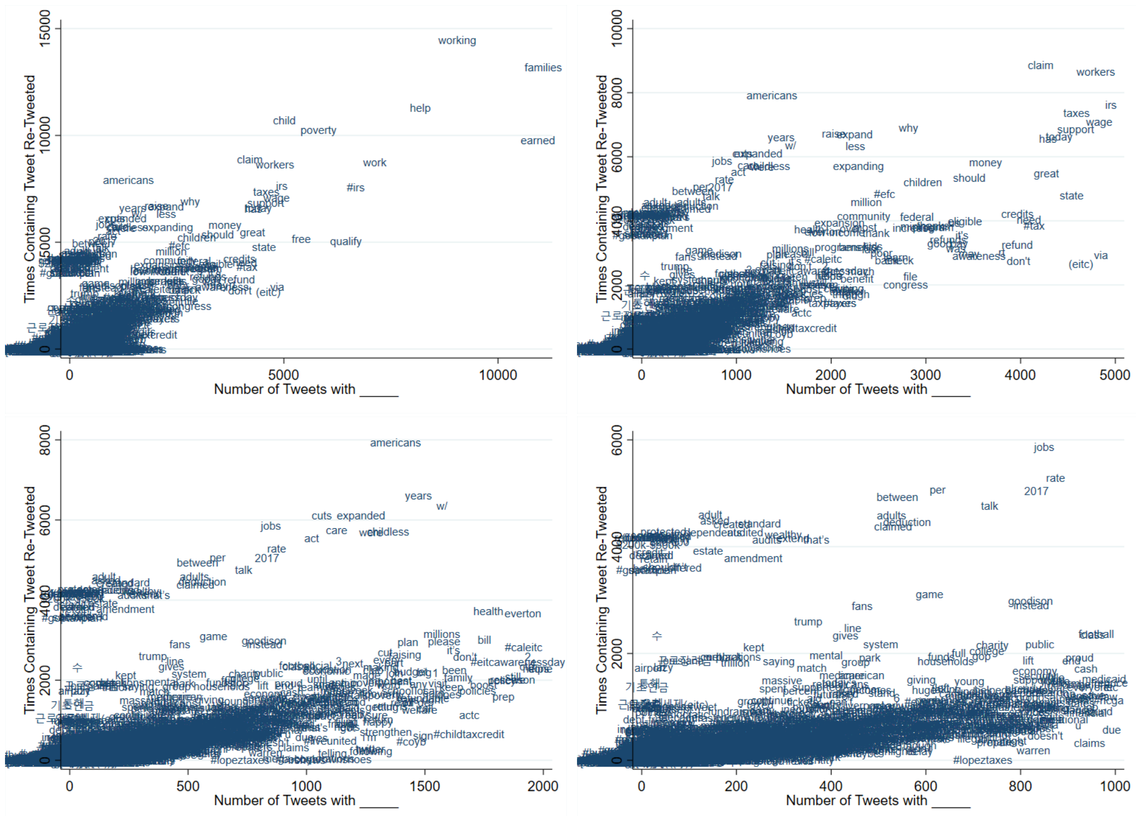

Figure A8: Re-Tweet Frequency by Word Count Frequency, January 2010-September 2019

Notes: Tweets since 2010 containing "EITC" were scraped from Twitter in September 2019. Tweets referencing the English Football club with handle @EITC were excluded. Scatter plot includes all words included in the tweet texts besides the 100 most common English words (e.g., "a", "about", etc.) and other common words like "&", "is", and "are". Words that appear less than 5 times are also omitted from the graph for clarity. Each panel restricts the number of tweets on the x-axis, to display less frequently tweeted words.

Source: Author's own calculations. Tweets posted between January 2010 and September 2019 taken from Twitter.

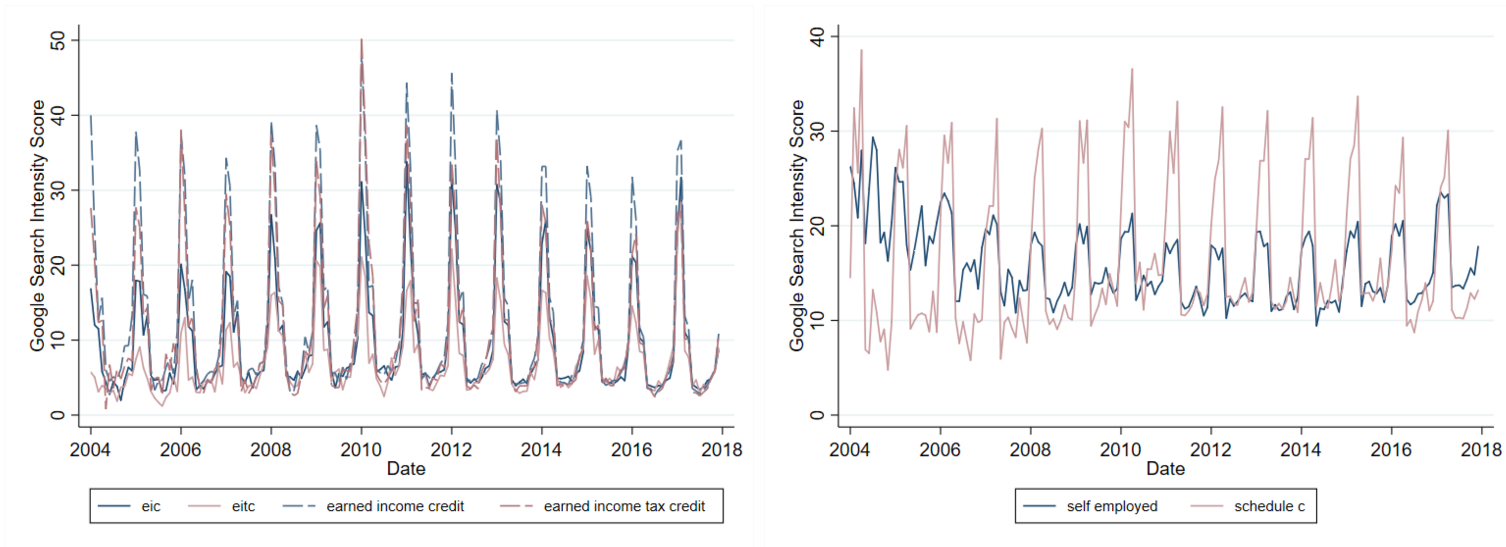

Figure A9: Google Search Intensity for EITC terms throughout the Year

Notes: Each figure represents the Google search index for the listed terms in every month, averaged across all designated market areas (DMA) in the county. For each DMA, the month with the highest search intensity is given a value of 100. Every other month is given a score between 0 and 100, proportional to the number of searches relative to the maximum. This measure is then averaged across all DMA for each year-month observation separately.

Source: Author's own calculations. Google search intensity taken from Google Trends at the designated market area level by year-month level.

Table A1: Direct Impact of State EITC on EITC Filing Behavior, Generalized Fixed Effects Specification

Generalized Fixed Effects	Percent of Returns with EITC			
Panel A. Levels	Any EITC	No Self Employment (No Schedule C, E, or F)	Self Employment (Schedule C, E, or F)	Percent of EITC Returns with Self Emp. (Schedule C, E, or F)
	(1)	(2)	(3)	(4)
Any State EITC	0.21	0.04	0.17	1.19*
	(0.24)	(0.25)	(0.18)	(0.63)
Dependent Mean	18.1	14.0	4.1	22.8
Observations	43,363	43,363	43,363	43,363
Generalized Fixed Effects	**Log Percent of Returns with EITC**			
Panel B. Logs	Any EITC	No Self Employment (No Schedule C, E, or F)	Self Employment (Schedule C, E, or F)	Log Percent of EITC Returns with Self Emp. (Schedule C, E, or F)
	(1)	(2)	(3)	(4)
Any State EITC	0.02	0.00	0.08***	0.06**
	(0.02)	(0.02)	(0.03)	(0.03)
Dependent Mean	18.1	14.0	4.1	22.8
Observations	43,363	43,363	43,347	43,347

Notes: Observation at the county by year level from 2000 to 2013. EITC claiming data obtained from the Brookings/IRS EITC interactive. Estimation controls for the county-level gender and race shares. Observations are weighted by the 2000 county population. County and year fixed effects are included, making this a comparison between counties in EITC expansion states and non-expansion states. Standard errors are corrected for clustering at the state level. <0.01 ***, p<0.05**, p<0.1*.

Table A2: Direct Impact of State EITC on Income Distribution of EITC Filers, Generalized Fixed Effects Specification

		Percent of EITC Returns by Income Bin								
	Average Refund	Under $5K	$5-10K	$10-15K	$15-20K	$20-25K	$25-30K	$30-35K	$35-40K	Over $40K
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Any State EITC	15.75	-0.30	0.16	0.28	0.02	-0.06	0.02	0.06	-0.02	-0.04
	(12.33)	(0.22)	(0.16)	(0.37)	(0.26)	(0.21)	(0.23)	(0.16)	(0.10)	(0.12)
Dependent Mean	1953.9	12.2	20.3	18.9	14.2	12.6	10.3	6.0	2.4	1.2
Observations	43,363	43,363	43,363	43,363	43,363	43,363	43,363	43,363	43,363	43,363

Notes: Observation at the county by year level from 1999 to 2013. EITC claiming data obtained from the Brookings/IRS EITC interactive. Estimation controls for the county-level gender and race shares. Observations are weighted by the 2000 county population. County and year fixed effects are included, making this a comparison between counties in EITC expansion states and non-expansion states. Standard errors are corrected for clustering at the state level. <0.01 ***, p<0.05**, p<0.1*.

Table A3: Direct Impact of State EITC on EITC Filing Behavior, Alternative Measures

Percent in EITC Range	Percent of Returns with EITC			
Panel A.	Any EITC	No Self Employment (No Schedule C, E, or F)	Self Employment (Schedule C, E, or F)	Percent of EITC Returns with Self Emp. (Schedule C, E, or F)
	(1)	(2)	(3)	(4)
Any State EITC*Percent EITC HH in 2000 with Income Below 40K	0.03***	0.00	0.02***	0.01
	(0.01)	(0.01)	(0.01)	(0.03)
Dependent Mean	18.1	14.0	4.1	22.9
Observations	43,232	43,232	43,232	43,232
Bunching Measure	Percent of Returns with EITC			
Panel B.	Any EITC	No Self Employment (No Schedule C, E, or F)	Self Employment (Schedule C, E, or F)	Percent of EITC Returns with Self Emp. (Schedule C, E, or F)
	(1)	(2)	(3)	(4)
Any State EITC*Percent EITC HH in 2000 with Income 5K-15K	0.01	-0.08***	0.09***	0.38***
	(0.02)	(0.02)	(0.02)	(0.08)
Dependent Mean	17.9	14.0	4.1	22.9
Observations	46,437	43,349	43,349	43,349

Notes: Observation at the county by year level from 2000 to 2013. EITC claiming data obtained from the Brookings/IRS EITC interactive. Estimation controls for the county-level gender and race shares. Observations are weighted by the 2000 county population. County and state by year fixed effects are included, making this a comparison between counties in the same state with more or less density below $40K in 2000 or "bunching" in 2000. These specification mirrors the specification in equation (2), but the intensity of treatment is captured by (1) the percent of returns in 2000 with income below $40K or (2) the percent of returns that fall in the part of the income distribution that corresponds to the peak of the EITC ($5-15K). Standard errors are corrected for clustering at the state level. <0.01 ***, p<0.05**, p<0.1*.

Table A4: Direct Impact of State EITC on Income Distribution of EITC Filers, Alternative Measures

		Percent of EITC Returns by Income Bin								
	Average Refund	Under $5K	$5-10K	$10-15K	$15-20K	$20-25K	$25-30K	$30-35K	$35-40K	Over $40K
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Any State EITC*Percent HH	2.66***	-0.04***	0.01	-0.00	0.02***	0.03***	0.04***	-0.02***	-0.03***	-0.03***
in 2000 with Income Below 40K	(0.58)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
Dependent Mean	1930.6	12.2	20.3	18.9	14.2	12.6	10.3	6.0	2.4	1.2
Observations	46,320	43,232	43,232	43,232	43,232	43,232	43,232	43,232	43,232	43,232
		Percent of EITC Returns by Income Bin								
	Average Refund	Under $5K	$5-10K	$10-15K	$15-20K	$20-25K	$25-30K	$30-35K	$35-40K	Over $40K
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Any State EITC*Share EITC HH	5.57**	-0.16***	-0.05***	0.07	0.10***	0.19***	0.16***	-0.03	-0.08**	-0.10***
in 2000 with Income 5K-15K	(2.25)	(0.04)	(0.02)	(0.06)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)
Dependent Mean	1954.0	12.2	20.3	18.9	14.2	12.6	10.3	6.0	2.4	1.2
Observations	43,349	43,349	43,349	43,349	43,349	43,349	43,349	43,349	43,349	43,349

Notes: Observation at the county by year level from 1999 to 2013. EITC claiming data obtained from the Brookings/IRS EITC interactive. Estimation controls for the county-level gender and race shares. Observations are weighted by the 2000 county population. County and state by year fixed effects are included, making this a comparison between counties in the same state with more or less density below $40K in 2000 or "bunching" in 2000. These specification mirrors the specification in equation (2), but the intensity of treatment is captured by (1) the percent of returns in 2000 with income below $40K or (2) the percent of returns that fall in the part of the income distribution that corresponds to the peak of the EITC ($5-15K). Standard errors are corrected for clustering at the state level. <0.01 ***, p<0.05**, p<0.1*.

Table A5: Robustness to Specification

	Baseline (1)	Trim top 1% (2)	No Controls (3)	County Trends (4)	Year F.E. (5)	DMA by Year F.E. (6)	Exclude Border Cnty. (7)	Outcome in Logs (8)	Un-weighted (9)	Un-weighted, in Logs (10)
	Percent of Returns with EITC, Self Employment									
Out-of-State Friends per person Exposed to State EITC (Std.)	0.36**	0.43***	0.40**	0.09**	0.44***	0.49***	0.58***	0.11***	0.06	0.04***
	(0.14)	(0.15)	(0.16)	(0.03)	(0.15)	(0.18)	(0.19)	(0.02)	(0.07)	(0.01)
Observations	43,349	43,166	43,349	43,349	43,363	43,265	37,133	43,333	43,361	43,345
	Percent of Returns with EITC, No Self Employment									
Out-of-State Friends per person Exposed to State EITC (Std.)	-0.35**	-0.31*	-0.25	-0.33***	-0.49***	-0.58***	-0.44*	-0.05***	0.08	-0.00
	(0.15)	(0.16)	(0.19)	(0.11)	(0.14)	(0.13)	(0.24)	(0.01)	(0.08)	(0.01)
Observations	43,349	43,166	43,349	43,349	43,363	43,265	37,133	43,349	43,361	43,361

Notes: Observation at the county by year level from 2000 to 2013. Out-of-State Friendship data constructed from the Social Connectedness Index (Bailey et al., 2018b). County migration data obtained from the IRS SOI county-to-county flows. Estimation controls for the county-level gender and race shares. Observations are weighted by the 2000 county population. County fixed effects are included to control for time invariant county characteristics. State by year fixed effects are included, making this a within state comparison. Standard errors are corrected for clustering at the state level. <0.01 ***, $p<0.05$**, $p<0.1$*.

Table A6: Impact of Out-of-State Friend Exposure to State EITC on EITC Filing Behavior, Exploiting County Network and 1999 EITC Rate

	Percent of Returns with EITC			
	Any EITC	No Self Employment (No Schedule C, E, or F)	Self Employment (Schedule C, E, or F)	Percent of EITC Returns with Self Emp. (Schedule C, E, or F)
	(1)	(2)	(3)	(4)
Out-of-State Friends per person	0.03	-0.31**	0.33***	1.92***
Exposed to State EITC (Standardized	(0.10)	(0.12)	(0.11)	(0.46)
and Weighted by 1999 EITC Rate)				
Dependent Mean	18.1	14.0	4.1	22.9
Observations	43,349	43,349	43,349	43,349

Notes: Observation at the county by year level from 2000 to 2013. Out-of-State Friendship data constructed from the Social Connectedness Index (Bailey et al., 2018b). The variable of interest is similar to that from equation (3), except the number of county-to-county friends is multiplied by the share claiming the EITC in 1999. This more closely matches the variation exploring the direct effects. The identifying assumption is more conservative, as we are now comparing claiming behavior between counties that have high linkages in high EITC counties of expansion states relative to counties with fewer linkages. The pattern of results also holds when multiplying by the share of households in EITC range and the share of EITC households bunching. EITC claiming data obtained from the Brookings/IRS EITC interactive. Estimation controls for the county-level gender and race shares. Observations are weighted by the 2000 county population. County fixed effects are included to control for time invariant county characteristics. State by year fixed effects are included, making this a within state comparison. Standard errors are corrected for clustering at the state level. <0.01 ***, p<0.05**, p<0.1*.

Table A7: Heterogeneity of Impact of Out-of-State Friend Exposure to State EITC on Family-level Reported Self-Employment (ACS 2005-2017)

	Self-Employment Rate				
	Some College or Less (1)	College Degree (2)	EITC Eligible Children (3)	No EITC Eligible Children (4)	Single Mother in Family Unit (5)
Out-of-State Friends per person Exposed to State EITC (Standardized)	0.24* (0.14)	0.29* (0.17)	0.07 (0.45)	0.12 (0.07)	0.38* (0.20)
Dependent Mean	13.1	17.1	17.7	13.1	10.4
Observations	11,479	11,479	11,479	11,479	11,479

Notes: Observation at the commuting zone/state level from the ACS between 2005 and 2017. Out-of-State Friendship data constructed from the Social Connectedness Index (Bailey et al., 2018b), and aggregated up to the commuting zone/state level. This measure is standardized by the same mean and standard deviation as above, so that coefficients can be directly compared. Estimation controls for area level age group shares, marital status shares, race and ethnicity group shares, education group shares, share of family units with a single mother, and the share of family units with one, two, or more EITC eligible children. Observations are collapsed from the family unit level using the household weights provided in the ACS, where individuals are probabilistically assigned to commuting zones as in (Autor et al., 2013)but the state of residence is maintained resulting in commuting zone by state level geographies. Commuting zone/state geographies are then weighted by 2010 population. Commuting zone/state fixed effects are included to control for time invariant geographic characteristics. State by year fixed effects are included, making this a within state comparison. Standard errors are corrected for clustering at the state level. <0.01 ***, p<0.05**, p<0.1*.

Table A8: Impact of Out-of-State Friend Exposure to State EITC on Self Filing

	Percent of Returns with EITC and Self Filed	
	Level	Logs
	(1)	(2)
Out-of-State Friends per person	0.50**	0.01
Exposed to State EITC (Standardized)	(0.21)	(0.02)
Dependent Mean	5.8	1.7
Observations	43,349	43,228

Notes: Observation at the county by year level from 2000 to 2013. Out-of-State Friendship data constructed from the Social Connectedness Index (Bailey et al., 2018b). EITC claiming data obtained from the Brookings/IRS EITC interactive. Observations are weighted by the 2000 county population. County fixed effects are included to control for time invariant county characteristics. State by year fixed effects are included, making this a within state comparison. Standard errors are corrected for clustering at the state level. <0.01 ***, p<0.05**, p<0.1*.

Table A9: Impacts of Out-of-State Friend Exposure by Concentration of EITC Income-Eligible Population

	Percent of Returns with EITC			
	Any EITC	No Self Employment (No Schedule C, E, or F)	Self Employment (Schedule C, E, or F)	Percent of EITC Returns with Self Emp. (Schedule C, E, or F)
	(1)	(2)	(3)	(4)
Out-of-State Friends per person	0.02	0.03	-0.01	1.31**
Exposed to State EITC (Standardized)	(0.17)	(0.15)	(0.17)	(0.53)
Out-of-State Friends Exposed to EITC	-0.02	-0.72***	0.70***	1.90***
*1999 EITC Share Above State Median	(0.11)	(0.15)	(0.14)	(0.50)
Dependent Mean	18.1	14.0	4.1	22.9
Observations	43,349	43,349	43,349	43,349

Notes: Observation at the county by year level from 2000 to 2013. Out-of-State Friendship data constructed from the Social Connectedness Index (Bailey et al., 2018b). EITC claiming data obtained from the Brookings/IRS EITC interactive. Estimation controls for the county-level gender and race shares. Observations are weighted by the 2000 county population. County fixed effects are included to control for time invariant county characteristics. State by year fixed effects are included, making this a within state comparison. Standard errors are corrected for clustering at the state level. <0.01 ***, $p<0.05$**, $p<0.1$*.

Table A10: Impact of Out-of-State Friend Exposure on Income Distribution of EITC Filers by Concentration of EITC Income-Eligible Population

		Percent of EITC Returns by Income Bin								
	Average Refund	Under \$5K	\$5-10K	\$10-15K	\$15-20K	\$20-25K	\$25-30K	\$30-35K	\$35-40K	Over \$40K
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Out-of-State Friends per person	-31.92***	-0.06	0.40***	1.17***	-0.43***	-0.33***	-0.07	0.16**	-0.02	-0.19**
Exposed to State EITC (Standardized)	(7.68)	(0.09)	(0.13)	(0.13)	(0.13)	(0.09)	(0.07)	(0.07)	(0.09)	(0.09)
Out-of-State Friends Exposed to EITC	61.27***	-0.59***	-0.12	-0.39***	0.84***	0.73***	0.74***	-0.15**	-0.60***	-0.69***
*1999 EITC Share Above State Median	(7.88)	(0.10)	(0.11)	(0.12)	(0.11)	(0.08)	(0.08)	(0.06)	(0.07)	(0.08)
Dependent Mean	1954.0	12.2	20.3	18.9	14.2	12.6	10.3	6.0	2.4	1.2
Observations	43,349	43,349	43,349	43,349	43,349	43,349	43,349	43,349	43,349	43,349

Notes: Observation at the county by year level from 2000 to 2013. Out-of-State Friendship data constructed from the Social Connectedness Index (Bailey et al., 2018b). EITC claiming data obtained from the Brookings/IRS EITC interactive. Estimation controls for the county-level gender and race shares. Observations are weighted by the 2000 county population. County fixed effects are included to control for time invariant county characteristics. State by year fixed effects are included, making this a within state comparison. Standard errors are corrected for clustering at the state level. <0.01 ***, p<0.05**, p<0.1*.

Table A11: Predictive Power of Distance and Interstate Linkage on Friendship Links

	Scaled Number of Friends		
	(1)	(2)	(3)
Distance (1000s miles)	-789.4***		-503.9***
	(114.1)		(79.1)
Same Interstate		17675.1***	32673.1***
		(2013.7)	(3804.5)
Distance			-23271.1***
*Same Interstate			(3412.6)
F-statistic	47.8	77.0	25.5
Dependent Mean	1027.4	1027.4	1027.4
Observations	9,533,148	9,533,148	9,533,148

Notes: One observation for each county-to-county pair included. Scaled number of friends constructed from the Social Connectedness Index (Bailey et al., 2018b). County fixed effects are included to control for county specific characteristics. Standard errors are corrected for clustering at the state level. <0.01 ***, $p<0.05$**, $p<0.1$*.

Table A12: Impact of Out-of-State Friend Exposure to State EITC and the Potential Role of In-Migration

	Individual In Migration Rate - EITC States	Percent of Returns with EITC		
		Any EITC	No Self Employment (No Schedule C, E, or F)	Self Employment (Schedule C, E, or F)
	(1)	(2)	(3)	(4)
Out-of-State Friends per person	0.40***	0.05	-0.30*	0.35**
Exposed to State EITC (Standardized)	(0.05)	(0.14)	(0.15)	(0.14)
Dependent Mean	0.4	18.1	14.0	4.1
Observations	43,386	43,349	43,349	43,349

Notes: Observation at the county by year level from 1999 to 2013. Out-of-State Friendship data constructed from the Social Connectedness Index (Bailey et al., 2018b). County migration data obtained from the IRS SOI county-to-county flows and captures in migration from different states. Columns (2)-(4) include household out-of-state in-migration as a control. Estimation controls for the county-level gender and race shares. Observations are weighted by the 2000 county population. County fixed effects are included to control for time invariant county characteristics. State by year fixed effects are included, making this a within state comparison. Standard errors are corrected for clustering at the state level. <0.01 ***, p<0.05**, p<0.1*

Table A13: Impacts Before and After Facebook Open to Everyone

	Percent of Returns with EITC			
	Any EITC	No Self Employment (No Schedule C, E, or F)	Self Employment (Schedule C, E, or F)	Percent of EITC Returns with Self Emp. (Schedule C, E, or F)
	(1)	(2)	(3)	(4)
		Before Facebook Began (Pre 2004)		
Out-of-State Friends per person Exposed to State EITC (Standardized)	0.61*	0.51***	0.10	-0.88**
	(0.31)	(0.18)	(0.17)	(0.35)
Dependent Mean	16.3	13.1	3.2	19.9
Observations	12,379	12,379	12,379	12,379
		After Facebook Began (Post 2004)		
Out-of-State Friends per person Exposed to State EITC (Standardized)	-0.06	-0.41***	0.35***	2.25***
	(0.11)	(0.10)	(0.12)	(0.43)
Dependent Mean	18.8	14.3	4.5	24.1
Observations	30,970	30,970	30,970	30,970

Notes: Observation at the county by year level from 2000 to 2013. Out-of-State Friendship data constructed from the Social Connectedness Index (Bailey et al., 2018b). EITC claiming data obtained from the Brookings/IRS EITC interactive. Estimation controls for the county-level gender and race shares. Observations are weighted by the 2000 county population. County fixed effects are included to control for time invariant county characteristics. State by year fixed effects are included, making this a within state comparison. Standard errors are corrected for clustering at the state level. <0.01 ***, p<0.05**, p<0.1*.