The Social Impact of the Earned Income Tax Credit

Eduardo Fajnzylber*
eduardo.fajnzylber@hec.ca

December 2003

Abstract

Currently, the Earned Income Tax Credit (EITC) is the most expensive cash-assistance program in the U.S., while also representing one of the main changes in recent welfare policy by providing positive incentives for work to families with children. Both its expansion in the early 1990s, and the more recent welfare reform legislation suggest there is strong concern regarding the impact of assistance programs on labor force participation. While most EITC research has focused on the direct effect of the program on labor force participation, little work has been done on the indirect effects: encouraging one woman to work may have positive spillovers on other eligible women. In this paper, we estimate the social impact of the EITC by comparing the labor force participation decisions of single mothers living in neighborhoods with different participation rates.

In the presence of positive spillovers, previous estimates of the impact of the EITC would underestimate the social impact of the program. The size of the "social multiplier" will depend on the intensity of the social interactions. We overcome the main identification issues involved in estimating this type of effect, by exploiting the 1993 expansion of the EITC to construct an instrument and a short panel of single mothers in California to allow for individual fixed effects. Our findings suggest that the social impact of the EITC expansion was an increase in the range of 2.2 to 3.2 percentage points in the labor force participation of single mothers, which can be decomposed into a (precisely estimated) 1.1% private effect and a 1.1% - 2.1% spillover effect.

JEL classification codes: J2, I3, Z13

*HEC Montréal, 3000 chemin de la Côte-Sainte-Catherine, Montréal (Québec), Canada H3T 2A7. I am deeply indebted to V. Joseph Hotz and Ekaterini Kyriazidou for their helpful advice and direction. In addition, I would like to thank Oscar Mitnik, Paul Devereux, Jeffrey Grogger, Enrico Moretti, Charles Mullin, Jinyong Hahn, Janet Currie, Robert Clarke and participants at the Labor Proseminar and Workshop at UCLA for their helpful comments. The errors are all mine. Financial support from the UC Institute for Labor and Employment and the UCLA Dissertation Year Fellowship are gratefully acknowledged.

1 Introduction

One of the main changes in social policy in recent years has been the expansion of the Earned Income Tax Credit (EITC) in the early 1990s. The EITC is a pro-work social program for families with children. It provides zero benefits to non-working families and as the family enters the labor market, the benefits increase with earnings up to a certain point.[1] Several studies have documented that the EITC could be partly responsible for the substantial increase in the labor force participation of single mothers during the 1990s.[2] The apparent success of the Federal EITC has led several States to implement their own versions of the program.[3] Even though the employment effects seem attractive, the program costs in terms of foregone tax revenues are an important consideration.[4]

Previous studies have used individual level data to estimate the differential impact of the EITC on families with different characteristics (in this case, family size). We refer to these effects as the *private effects* of the EITC in the sense that they can be attributed to the changes in the incentives faced by each individual family, regardless of the decisions taken by other individuals. However, these studies have not taken into account the indirect effects of the program: encouraging one woman to work may have positive spillovers on other women in the same neighborhood. We refer to these indirect effects as *spillover effects*. These effects, if positive and significant, can generate a *social multiplier* that would amplify the private effect of the EITC.[5] We refer to the overall effect (the sum of the private effect and the spillover effect) as the *social impact* of the EITC.[6] The magnitude of this social multiplier depends on the intensity of the *social interactions* between individuals, i.e. to what extent the fraction of working families in a neighborhood affect an individual's labor force participation decision.

In this article, we estimate the magnitude of the spillover effect of the EITC, by comparing the labor force participation decision of single mothers living in neighborhoods with different proportions of participating individuals between 1993 and 1994, the period of the largest EITC expansion. We use this estimate to construct a measure of the social impact of the program.

Traditional economic models have assumed that decision making is based on preferences over different goods and constraints imposed by the resources available to an individual (or a family). The total demand for a product is obtained by simple aggregation of individual choices. However, other social sciences, particularly sociology or antropology, have long believed in the importance of the social environment on the decisions made by individuals. By measuring separately the private

[1] See Appendix 1 for a brief description of the EITC. Hotz and Scholz (2000) provide an extensive survey of the research addressing the EITC and its impact on economic and behavioral outcomes.

[2] See for example Eissa and Liebman (1996), Dickert, Houser, and Scholz (1995), Meyer and Rosenbaum (1999), Ellwood (2000), Grogger (2003a), Grogger (2003b), and Hotz, Mullin, and Scholz (2001).

[3] By 2000, 15 states had implemented some form of State EITC. The most common variant is simply to allow a family to claim a fixed percentage (in general, 10%) of the Federal EITC.

[4] In 1999, for instance, the total Federal expenditure on the EITC was $31.9 billion.

[5] See Glaeser, Sacerdote, and Scheinkman (2002).

[6] The concepts of private and social are relative to the type of variation used to identify the corresponding effects. If , for example, the EITC only affected some individuals within a family but not all of them, using family level data would capture the "family social effect" in the sense that it would embed the individual level effects and any type of spillover within the family.

and spillover effects of changes in the incentive structure surrounding the EITC expansion, we directly test the hypothesis that the environment of an individual, in this case the neighborhood, can have a direct influence on her decisions[7].

From a policy perspective, and given the high cost of the EITC, it is important to quantify both the private and the spillover effects of the program. The private effect represents the result of increasing the incentives to work by subsidizing the initial wage perceived by a low income worker, independent of the location of the individual. Predictions on the direct effect of further raising these incentives, for example by introducing a State EITC, could be obtained from these estimates.

On the other hand, if the spillover effects are important, it might be a more cost-effective approach to implement policies that act directly on the social interactions. For example a policy designed to increase the flow of information about job opportunities or the benefits of the EITC, especially in low participation neighborhoods, could have large effects on labor force participation, at a relatively low cost.

To estimate the pure effect of interacting individuals, we need to address the identification problems associated with trying to explain individual behavior with mean behavior in the individual's reference group (Manski 's "reflection problem"[8]). As the outcome of individual B determines the outcome of individual A in the same neighborhood (and viceversa), we have a standard *simultaneity problem.* Depending on the nature of the social interactions (endogenous or exogenous), the reduced form will not identify all the structural coefficients.[9] The second problem, *endogeneity caused by correlated unobservables*,[10] occurs when individual characteristics (specially the unobserved ones) are correlated among individuals belonging to a certain group. This case is of particular concern when individuals self-select into groups (*endogenous membership*). Even in the absence of social interactions, similar individuals will tend to make similar choices, causing a spurious correlation between individual and group behavior. The third problem, *unobserved group characteristics,* originates from the difficulty in controlling for all neighborhood characteristics that affect the outcome of interest. Those omitted characteristics could generate a correlation in the decisions taken by individuals in the same group that is unrelated to any form of social interaction.

Previous studies have dealt in different ways with the reflection and unobserved neighborhood characteristic problems. We discuss some of these approaches in section 2. The correlated unobservables problem has remained a difficult issue to control directly[11].

[7]The experimental evaluation literature on social interactions has found evidence of social effects in retirement plan decisions (Duflo and Saez (2002)), and academic effort and social groups (Sacerdote (2001)). Marmaros and Sacerdote (2002) also find evidence of peer effects in employment outcomes of college graduates from other students living in the same dorm hallway. Katz, Kling and Liebman (2000) analyze data from a randomized housing voucher program and find that low income families who moved to low-poverty neighborhoods experienced improvements in different measures of well being but no significant short-run effects on employment, earnings or welfare receipt. Ludwig, Duncan and Pinkston (2000) find some evidence that moving to a low-poverty neighborhood might reduce welfare participation and increase employment.

[8]See Manski (1993).

[9]Manski (1993) calls *endogenous* the effects of average behavior on individuals decisions and *exogenous*, the effect of mean exogenous characteristics of the individuals in a person's reference group.

[10]See Manski (1993) or Moffitt (forthcoming).

[11]An exception is Moretti (forthcoming), who includes individual*city fixed effect to account for sorting of high

We use the EITC expansion between 1993 and 1994 that differentially affected families with one child versus families with two or more children to overcome the unobserved group component and reflection problems[12,13]. The average fraction of one child versus two-or-more children families in a neighborhood (interacted with a 1994 year dummy variable) can be used as an instrument to isolate the impact of the average labor force participation in the neighborhood on individual behavior. The idea is that a woman living in a neighborhood with a high fraction of two-or-more-children is more likely to be "exposed" to a high increase in the average labor force participation rate. However, the EITC expansion is not a perfect experiment. The intensity of the treatment is a function of family size, which might be correlated with unobserved determinants of labor force participation. Self selection into neighborhoods will potentially cause these unobserved factors to correlate among individuals. We need a strategy to deal with the endogenous membership problem.

Given that endogenous membership is essentially a sample selection problem, we use an individual fixed effect approach on a two-period (1993 and 1994) longitudinal data set composed of single mothers participating on welfare at some point between 1992 and 1995.[14] Assuming that, in the short run, the distribution of individuals within neighborhoods remains stationary, the correlated unobservables problem caused by endogenous membership is "differenced out" by the individual fixed effect for those individuals that did not change their zip code. Assuming that family size is uncorrelated with unobserved determinants of employment (conditional on the individual fixed effect), the average fraction of one versus two-or-more children families in a neighborhood will be a valid instrument for the changes in the average labor force participation. Additional controls for local economic conditions (at the county level), will eliminate most of the remaining endogeneity problems caused by unobserved neighborhood characteristics. We also estimate models that more robustly control for changes in local economic conditions or changes in the work emphasis of welfare offices of different counties, by including county*time fixed effects.

We use administrative data from the state of California, which provides the universe of women receiving welfare benefits in California, at some point between 1992 and 1995. Our neighborhood definition corresponds to the zip code of residence of these women and their families.

We estimate a private effect of the EITC of the same magnitude as found in previous studies. Between 1993 and 1994, single mothers with two or more children experienced an increase in their labor force participation by approximately 1.6 percentage points more than families with only one child. As we show, this estimate is robust to the inclusion of the social interaction effect and to different sample definitions. The aggregate private impact of the program was to raise the overall labor force participation among low income mothers in California by 1.1 percentage points.

ability workers into cities.

[12] This idea was originally mentioned in Moffitt (forthcoming), who suggested that "...partial-population experiments in which only a portion of the individuals within each group are given a treatment..." could provide an identification strategy that didn't involve experimental manipulation of group membership (Page 20).

[13] The maximum benefit for families with two children rose from $1511 in 1993 to $3110 in 1995, while families with one child increased their maximum benefit from $1434 to $2094 in the same period.

[14] This strategy is proposed in Kyriazidou (1997). Brock and Durlauf (2001) suggest that panel data, under appropriate conditions, could help solve the self-selection problem in interactions-based environments (page 3337).

The estimated magnitude of the social interaction coefficient ranges from 0.46 to 0.63, which implies that if a woman was living in a neighborhood that increased its average labor force participation from 50% to 60% between 1993 and 1994, her likelihood of working would increase by 4.6 to 6.3 percentage points. Our combined effect, the aggregate social impact of the EITC between 1993 and 1994, ranges from 2.2 to 3.2 percentage points. The spillover effects are, however, less precisely estimated.

The remainder of the paper is structured as follows: in section 2, we present the basic specification that we will estimate in this paper, laying out the identification problems and the strategy we will follow to overcome them. In section 3 we present the data. In section 4, we present and discuss our results and robustness analyses and in section 5 we conclude.

Appendix A presents a brief summary of the main characteristics of the EITC program, Appendix B discusses the assumptions for identification of the linear social interactions model with individual fixed effects and instrumental variables and in Appendix C, we present a simple model of location and work decisions that illustrates how the endogenous sorting into neighborhoods can generate a correlation between individual and neighborhood averages, even in the absence of social interaction effects.

2 Empirical identification strategy

As a benchmark, we will estimate the private effect of the EITC using a similar model to the one used in the evaluation literature on the EITC, which exploits the differential treatment of the program for families with one child and families with two children and how this differential treatment was expanded between 1993 and 1994. This essentially corresponds to a difference in difference model including individual level characteristics as controls. The model is estimated for a sample of single mothers with children.

$$\begin{aligned} LFP_{igt} &= \alpha + \gamma \left(KIDS_{it}^{2+} * YEAR_{it}^{1994} \right) + \delta_1 KIDS_{it}^{2+} + \delta_2 YEAR_{it}^{1994} \\ &\quad + \delta_3 AGE_{it} + \theta' LEC_{gt} + \varepsilon_{it} \qquad (1) \\ t &= 1993, 1994 \\ i &= 1..N \end{aligned}$$

where LFP_{igt} corresponds to the labor force participation of woman i who lives in neighborhood g in time t, $YEAR_{it}^{1994}$ is a dummy variable equal to 1 if the observation corresponds to 1994 and zero otherwise, $KIDS_{it}^{2+}$ is a dummy variable equal to 1 if the woman has 2 or more children and 0 if the woman has only one child, AGE_{it} is the age of the mother at the end of year t, AGE_{it}^2 is the squared age of the mother, and LEC_{gt} is a measure of Local Economic Conditions in the area (county) where individual i resides in period t.

Under reasonable conditions, the ols estimate γ will consistently estimate the private effect of the EITC.[15]

To simplify notation, we will rewrite equation (1) as:

$$Y_{igt} = \alpha + \gamma EITC_{it} + \delta' X_{it} + \theta' Z_{gt} + \varepsilon_{igt}$$

where Y_{igt} is the labor force participation decision, $EITC_{it}$ corresponds to the interaction term, X_{it} is a vector of remaining individual characteristics and Z_{gt} represents neighborhood characteristics (observed).

Now, let's introduce the spillover effect of the EITC. The simplest way to model such effects is to include the average participation among individuals living in the same area as a regressor.[16]

$$Y_{it} = \alpha + \beta E\left[Y_{it}|G_{it}\right] + \gamma EITC_{it} + \delta' X_{it} + \theta' Z_{gt} + \left(\mu_{gt} + E\left[\varepsilon_{igt}|G_{it}\right] + \nu_{igt}\right) \qquad (2)$$

where G_{it} represents the neighborhood where individual i resides in period t and $E\left[Y_{it}|G_{it}\right]$ corresponds to the population mean labor force participation in neighborhood G_{it}.

Notice that we have assumed that the average $EITC$ variable in the neighborhood does not directly affect individual decisions. In this case, this is a reasonable assumption to the extent that the EITC expansion was an unanticipated event and that in the short run (1 year), neighborhoods are not likely to change in response to the policy.

We decomposed the error term in three terms, to reflect the main identification issues associated with the estimation of social interactions models:

- The term μ_{gt} represents *unobserved neighborhood characteristics.* As these components will affect each individual in the neighborhood in a similar way, it will be correlated with average participation in the neighborhood.

- The second term, $E\left[\varepsilon_{igt}|G_{it}\right]$, represents the problem of *correlated unobservables*; Individuals who **chose** to live in the same neighborhood (who have the same G_{it}) will potentially share common unobserved determinants of labor force participation. If not accounted, the existence of this term will generate a correlation in the decisions of individuals, even in the absence of social interactions. In appendix C, we present a simple location/work model to illustrate this issue.

- Finally, ν_{igt} is an error term, assumed uncorrelated with the regressors in the equation and the determinants of the location decision.

In practice, $E\left[Y_{it}|G_{it}\right]$ is replaced by the sample mean $\left(\frac{1}{N_{G_{it}}}\sum_{G_{jt}=G_{it}} Y_{jt}\right)$. This will generate a standard *simultaneity problem*: If individuals A and B live in the same neighborhood, their LFP decisions will enter each other's equation.

[15] See Hotz and Scholz (2000) for a discussion of the conditions for identification in this model.

[16] This assumption, the most common in the empirical literature, can be justified from the linearized version of a model of information diffusion (also known as "contagion models"). The probability of meeting an informed individual is a function of the average participation in a neighborhood.

In order to deal with the simultaneity and the unobserved group component problems, some studies have used different measures of the network available to an individual in a particular neighborhood:

- Bertrand et al. (2002) use an interaction between "contact availability" (the proportion of individuals that speak the same language in the neighborhood) and "quality of contacts" (average welfare use of the language group in the country):

$$Y_{ijg} = \beta\left(CA_{jg} * \overline{Y_j}\right) + \delta' X_i + \gamma_g + \eta_j + \theta CA_{jg} + \nu_{igt}$$

  where j represents the language group and g the neighborhood. As the interaction term varies for individuals from different language groups in the same neighborhood, they are able to include a neighborhood fixed effect (γ_g) that will capture any unobserved neighborhood characteristic (that is not group specific).

- Aizer and Currie (2002) estimate models where networks are measured as the average outcome (utilization of publicly funded prenatal care) among women in the same neighborhood and race group during the 11 months prior to a birth:

$$Y_{ijgt} = \beta_1\left(CA_{jgt} * \overline{Y_{jg(t-1)}}\right) + \beta_2\overline{Y_{jg(t-1)}} + \beta_3\overline{Y_{(-j)g(t-1)}} + \delta' X_i + \gamma_{gt} + \nu_{igt}$$

  where j and $-j$ represent the "own" and "others" language groups, g represents the neighborhood and γ_{gt} is a neighborhood*time fixed effect that will capture time-varying unobserved group characteristics. This dynamic model, with lagged outcomes on the RHS of the equation, is no longer affected by the simultaneity problem.

To identify the coefficients of interest (β and γ), we will follow a different approach. We first obtain a reduced form equation by taking conditional expectations over both sides of equation (2) and solving for $E\left[Y_{it}|G_{it}\right]$:

$$\begin{aligned} E\left[Y_{it}|G_{it}\right] &= \frac{\alpha}{1-\beta} + \frac{\gamma}{1-\beta}E\left[EITC_{it}|G_{it}\right] + \frac{\delta' E\left[X_{it}|G_{it}\right]}{1-\beta} + \frac{\theta' Z_{gt}}{1-\beta} \\ &\quad + \frac{\mu_{gt} + E\left[\varepsilon_{igt}|G_{it}\right]}{1-\beta} \end{aligned} \tag{3}$$

This represents the unique Social Interaction equilibrium of the system. Estimating this equation, it is possible to obtain the Social impact of the EITC as the coefficient corresponding to $E\left[EITC_{it}|G_{it}\right]$.

$$\text{Social Impact of the EITC } = \frac{\gamma}{1-\beta} = \gamma + \frac{\beta\gamma}{1-\beta} \tag{4}$$

To decompose the effect into a private (γ) and a spillover effect ($\frac{\beta\gamma}{1-\beta}$), we need to go back to the individual level equation, where we can substitute $E\left[Y_{it}|G_{it}\right]$ from equation (3).

$$\begin{aligned} Y_{it} &= \frac{\alpha}{1-\beta}+\frac{\gamma\beta}{1-\beta}E\left[EITC_{it}|G_{it}\right]+\frac{\beta\delta' E\left[X_{it}|G_{it}\right]}{1-\beta}+\gamma EITC_{it}+\delta' X_{it}+\frac{\theta' Z_{gt}}{1-\beta} \\ &\quad +\left(\frac{\mu_{gt}+E\left[\varepsilon_{igt}|G_{it}\right]}{1-\beta}+\nu_{igt}\right) \end{aligned} \tag{5}$$

As we discuss in appendix B, assuming exogeneity of the observed characteristics (including $EITC_{it}$) and exogenous allocation into neighborhoods, it is possible to identify all the coefficients from this reduced form equation (5). Imposing the restrictions of the Social equilibrium (3) could even lead to overidentification. of the model. It is, however, difficult to justify the exogeneity assumptions. The individual observed characteristics (X_{it}) are likely to be correlated with determinants of location decisions or with unobserved neighborhood characteristics. Even the EITC variable (which is a function of family size) could be correlated with unobserved determinants of the location decision.

For these reasons, we will exploit the longitudinal nature of our data set by including an individual fixed effect and restricting our analysis to the subsample of individuals that didn't change neighborhoods. Instead of estimating the two reduced form equations, we will use a two stage least squares estimation procedure where the equilibrium equation will be estimated in the first stage and the predicted average LFP will replace $E\left[Y_{it}|G_{it}\right]$ in (2). As we use a two-year panel of individuals (and we focus on the individuals who didn't move between 1993 and 1994), we can express the estimating model in first differences, in the following system:

$$\begin{aligned} \text{First Stage} &: \quad \Delta\overline{Y_{gt}}=\pi_0+\pi_1\Delta\overline{EITC_{gt}}+\pi_2'\Delta X_{it}+\pi_3'\Delta Z_{gt}+\eta_{gt} \\ \text{Second Stage} &: \quad \Delta Y_{igt}=\alpha_0+\beta\Delta\overline{Y_{gt}}+\gamma\Delta EITC_{it}+\delta'\Delta X_{it}+\theta'\Delta Z_{gt}+\Delta\varepsilon_{igt} \end{aligned}$$

This procedure, under the assumptions discussed in Appendix B, will serve three purposes:

- Capture time invariant neighborhood characteristics,
- Turn the EITC variable (conditional on the individual fixed effect) into a valid instrument for time varying unobserved group characteristics,
- Eliminate the selectivity bias introduced by endogenous sorting into neighborhoods.

As we will see in section 4.4, we also present some estimates based on the individual level reduced form equations, assuming that the system is in a social interaction equilibrium. In section 4.5, we repeat both estimation strategies for a model with county*time fixed effects, which allows us to control more effectively for determinants of work that might be changing over time at the county level, including local economic conditions and the particular emphasis on work requirements of the different counties.

3 The data

Estimation of the proposed identification strategy requires a data set that contains precise geographical information (zip code) on a high number of individuals belonging to an homogeneous population, with at least two years of longitudinal data.

The data set used is based on confidential administrative data from the State of California. The main source corresponds to a series of yearly files from the MediCal Eligibility Data System (MEDS), which keeps monthly eligibility codes for every individual eligible for MediCal, the State health insurance program. As individuals participating on the AFDC program are automatically incorporated into the MEDS file, this could be considered as a census of all individuals receiving welfare assistance in California. The MEDS file provides monthly information on the eligibility codes and county of residence, zip code information every six months, and a limited number of demographic characteristics like date of birth, race and language.

We had access to these files for the period 1987-2001 but in order to justify some of the assumptions required for identification (that neighborhoods didn't change dramatically in response to the EITC expansion), the analysis is limited to the 1993-1994 period, where the biggest expansions of the EITC benefits took place.

As we are interested in a sample of single mothers, we constructed a sample of women with ages between 19 and 45, in 'cases' where no other person over 18 is present, where at least one person under 18 is present and the oldest person under 18 has a difference of at least 13 years with the adult.

One crucial aspect of this data is that family composition and geographical information are only available for those periods where the person was eligible for MediCal. Our base sample corresponds to all the women that were present at least once during the period 1992-1995 and we applied a simple algorithm to impute location and number of children for the years in which the person was not observed:

1. We impute family composition (number and ages of children) and geographic location (zip code) backwards. If, for example, the person was on MediCal during 1993 and 1995 but not in 1994, we impute the first zip code observed in 1995 as the location for 1994 and we roll back the ages of children present during the first month of 1995 to obtain imputed family structure for 1994.

2. If no information is available from the future (up to 1995), then we impute forward. If a woman was on MediCal during 1994 but not in 1995, we impute the zip code in 1995 from the last zip code observation in 1994. We also roll forward the ages of children from the last observation in 1994.

In order to construct labor force participation for these individuals, we used matched information from the California Unemployment Insurance administration, which corresponds to quarterly data on the earnings received by each person from any job covered by this program. This only excludes government jobs and self-employment. To avoid problems with incorrect social security numbers (that could result in non matches) or numbers that were being used by multiple individuals, we further restricted the sample to individuals with verified identifiers and with less than 12 employers in a year. We also discarded families with more than 6 children. Our basic measure of individual labor force participation is a dummy variable equal to one if annual earnings were higher than $200. Notice that this measure is never imputed because we observe earnings independently of MediCal eligibility.

We also used county information to construct measures of local economic conditions. The California Unemployment Insurance administration provides yearly aggregate measures of employment (number of individuals) and earnings by county for 10 different industrial sectors. Combining the employment data with an estimation of the population aged 18-65 (from the California Department of Finance), we constructed an employment to population ratio for each county*year observation.[17]

3.1 Summary statistics

In table 2, we present summary statistics of the individuals in our baseline sample. Each observation corresponds to a single mother with at least one child. Besides from the validation conditions mentioned in the previous section, the baseline sample corresponds to all those individuals that didn't change zip code between 1993 and 1994, who lived in zip codes with at least 101 individuals (in our sample) and that didn't change the number of children between 1993 and 1994. The restriction to a population of 'stayers' is motivated by our identification strategy; for this sample, the individual fixed effect will effectively 'difference out' the selectivity bias. The minimum sample size restriction was introduced to minimize the measurement error inherent in estimating population means with sample means. Finally, the restriction in fertility is a way of controlling for changes in the EITC eligibility group; a woman who goes from having one child (low benefit group for the EITC) to having two children (high EITC eligibility group) is much more likely to reduce her labor force participation during the first year of life of the newborn.

Our data set includes 142391 individuals distributed in 503 zip codes. Average labor force participation rose between 1993 and 1994 from 33.7% to 36.7%. Approximately 69.1 % of these families have more than 1 child, with an average of 2.2 children. The sample is also predominantly hispanic (42.8%) and white (28.8%) and the average age of mothers is almost 34. As can be seen from our measures of local economic conditions, the Californian economy was improving between 1993 and 1994, with both employment increasing and unemployment falling.

We can also see that both the fraction of women working in a neighborhood and the fraction of one v/s two-or-more children families have a reasonable range of variation: average LFP goes from 13.8% in some areas to 58% in others, while the fraction of 2^+ children families varies from 0.386

[17] Based on the same sources, we tried other measures of local economic conditions: we constructed employment to population ratios for the retail and services sectors (which are traditional sources of jobs for the low-income population), as well as average earnings per worker. As an alternative measure, we used county unemployment rates calculated by the Bureau of Labor Statistics. As these alternative measures did not signifficantly affect the coefficients of interest, we didn't present the results of these specifications in the final version of the article.

to 0.838.

In terms of our estimation strategy, more relevant than the average LFP in a zip code is the change in that variable between 1993 and 1994. We present the distribution of this change in figure 3 and the distribution of the fraction of 1 vs 2 or more children families in figure 4. From figure 3, we notice that most neighborhoods raised their average LFP between 0 and 13 percentage points. There is also a certain fraction of areas that decreased their average participation.

4 Results

4.1 The private effect of the EITC

As a benchmark model, we present in table 3 the results of estimating the private effect of the EITC using our sample of single mothers. The first column corresponds to a cross sectional model for 1994. After controlling for age and age squared, the labor force participation of women with more than one child is 2.7 percent lower than that of women with two children. So, in general, larger families tend to work less than families with only one child.

The second column presents the results for a model where data from 1993 and 1994 were pooled together but instead of including a fixed effect, we allowed for clustering at the individual level. From this model, we obtain our first estimate of the private effect of the EITC, given by the coefficient on the interaction between the 'More than 1 child' dummy variable and the 'Post 1993' dummy variable. The results suggest that women with 2 or more children raised their labor force participation between 1993 and 1994 by 1.6 % more than women with only one child. This is a very similar result to the ones found in the evaluation literature of the EITC.

The other coefficients have the expected signs: the direct effect of having more children tends to lower participation but this variable tends to increase with age at a decreasing rate. The coefficient on the employment to population ratio, however, has the wrong sign.[18]

Finally, the third column presents the equivalent to column two with individual fixed effects, with extremely similar results.

4.2 Private effects, spillovers and the social impact of the EITC

Our main results are presented in table 4, which includes the estimates from three different models that allow for social interactions by including the average labor force participation on the RHS of the equation.

The first column corresponds to the estimation of a model that includes a fixed effect (that should eliminate the selectivity bias) but that doesn't include an instrument for the average participation in the neighborhood. As we explained earlier, failing to solve the simultaneity and unobserved group components problems can result in a strong endogeneity problem that will tend

[18] We experimented with different controls for local economic conditions, with practically no effect on the coefficients of interest. We only present the results with the most general control, the employment to population ratio for all sectors.

to bias upwards the social interaction coefficient.[19] In this case, the estimate is practically equal to one and extremely significant. Notice that the estimated private effect (presented in the second row of the table) remained practically unchanged from the results in table 3. This suggests that if the object of interest of the evaluation of a program is to obtain the private effect (rather than the social effect), it sounds reasonable to perform an individual level evaluation without having to worry about the spillover effects. Put in other terms, the spillover effect cannot be viewed as an omitted variable problem that will bias the results of a private effect evaluation.

The second column presents the result of the opposite exercise, in which we use the instrument generated by the expansion of the EITC but we do not include family fixed effects.[20] This can interpreted as a model in which we assume exogenous membership and exogeneity of the included regressors (like in the first case treated in Appendix B). Once again, if individuals who sort into the same neighborhoods tend to have similar propensity to work (or to participate on welfare programs), failure to account for selectivity effects will tend to bias upwards the spillover effects. In this case, we estimate a social interaction coefficient of 0.753, which could be interpreted by saying that if a woman was living in a neighborhood that increased its average labor force participation from 50% to 60% between 1993 and 1994, her likelihood of working would increase by 7.53 percentage points.

The other coefficients are in line with previous results (including a private effect of the EITC of 1.5%).

The first results based on our identification strategy are presented in the third column of table 4. In this baseline model, we use both the instrument and the individual fixed effects. Our results suggest a private effect of the EITC of 1.6%, a spillover coefficient of 0.651 and a social impact for the EITC (given by the first stage coefficient presented in the bottom panel of column three) of 0.046. Multiplying this coefficient by the average fraction of 1 vs 2 or more children families in our sample (0.691), we obtain the overall social impact of the EITC for California, between 1993 and 1994 (standard errors in parenthesis):

$$\begin{aligned}
\text{Overall social Impact of the EITC} &= \widehat{\pi_1} * \overline{\overline{\Delta EITC}} = 0.046 * 0.691 = 0.032 \;\; (0.014) \\
\text{Overall private effect of the EITC} &= \widehat{\gamma} * \overline{\overline{\Delta EITC}} = 0.016 * 0.691 = 0.011 \;\; (0.002) \\
\text{Overall spillover effect of the EITC} &= 0.032 - 0.011 = 0.021 \;\; (0.014)
\end{aligned}$$

Allong with the number of observations and R-squared of both stages, we report both the R-squared associated with the average EITC variable (the excluded instrument). The associated R-squared (with the included regressors partialled out) is 0.01, which is in the range where one would question the weakness of the instrument.

[19] In the extreme case of running a regression of an outcome against the average outcome, without other controls, Manski (1993) showed that the ols coefficient will mechanically be equal to one, independently of the existence of social interactions.

[20] To make the results more comparable accross specifications, we used the exact same sample in each case, which includes the restriction that individuals do not change zip code between 1993 and 1994. This is a restriction imposed in conjunction with the fixed effect approach. The results from the second model are very similar when using the entire sample.

4.3 Sensitivity Analysis

We estimate the baseline specification varying two characteristics of the sample selection procedure: in the first one, we vary the minimum cell size per zip code (the minimum number of individuals that we allow in a particular zip code) and in the second one, we restrict the sample to individuals for which we have multiple observations from which to impute missing location or family composition. The baseline result of table 4 is reproduced in the third column of table 5.

The first sensitivity analysis is presented in the first three columns of table 5. In the first column, we allow for a minimum cell size of 26 individuals, which more than triplicates the number of zip codes for estimation. We obtain a smaller social interaction coefficient (0.574) and a slightly larger EITC coefficient. Similarly, when we allow for a minimum size of 51 individuals per zip code, we obtain a smaller coefficient than with the baseline sample.

As we mentioned earlier, our base sample includes all mothers who were eligible for MediCal at least once between 1992 and 1995 and location and family size is imputed for the missing years. This imputation algorithm requires stronger assumptions in cases where a person is observed only once during the four years. For this reason, we constructed two additional subsamples:

- Sample B: On MediCal during the periods (1993 and 1994) or (1992 and 1994) or (1993 and 1995) or (1992 and 1995).
- Sample C: On MediCal during the periods (1993 and 1994) or (1992 and 1994) or (1993 and 1995).

The idea of these samples is that we only keep individuals for which we have at least two pair of non consecutive observations to impute data for 1993 and 1994.

The results are presented in the last three columns of table 5. The social interaction coefficient have a similar magnitude (slightly smaller) than the baseline results but with larger standard errors. This is quite likely explained by the diminished number of zip codes available to identify the social effect in the first stage (the number of zip codes falls from 503 to 364 when going from the baseline to Sample B and from there to 354 zip codes in sample C).

Overall, our results seem robust to the choice of sample that we used, even though more stringent requirements diminish the effective sample size to levels where our coefficients lose statistical significance.

4.4 Reduced form estimates

The previous estimates were based on the traditional assumptions of instrumental variable estimators for the model in first differences: that the instrument $\left(\Delta\overline{EITC_{gt}}\right)$ is correlated with the endogenous variable $\left(\Delta\overline{Y_{gt}}\right)$ while uncorrelated with the error term in the main equation. The first condition is motivated by the fact that the individual EITC variable determines individual participation, so we would expect that the aggregate EITC should explain average participation in the neighborhood. We did not impose any restriction on the coefficients of either stage.

However, if we assume that the system has reached the social interaction equilibrium of equation (3), we can use the Reduced Form equation (5) to solve for the social interaction coefficient (β):

$$\Delta Y_{igt} = \pi_0 + \pi_1 \Delta \overline{EITC_{gt}} + \pi_2 \Delta EITC_{it} + \pi_3' \Delta X_{it} + \pi_4' \Delta Z_{gt} + \eta_{gt} \tag{6}$$

where

$$\begin{aligned} \beta &= \frac{\pi_1}{\pi_1 + \pi_2} \\ \gamma &= \pi_2 \end{aligned}$$

The results for this procedure are presented in table 6, where the upper panel corresponds to the coefficient from the reduced form equation (6) for the different samples defined in the previous section, and the lower panel presents the estimates of β using the previous formula.[21] Comparing these estimates with the first row of table 5, we can see that the two procedures give very similar results, with the reduced form coefficients being slightly higher than the 2SLS ones. This suggests that the social interaction equilibrium assumption seems to be a valid one.

4.5 Models with county*time fixed effects

In the previous specifications, we were controlling for local economic conditions by including a proxy variable for changes in the labor demand at the county level (employment to population ratio). As this measure, or any other, are potentially subject to criticisms, we introduce county*time fixed effect to account for any determinant of labor force participation that may be varying at the county level. This includes, for example, any changes in the emphasis towards work from the county welfare offices.

The results from these specification are presented in tables 7 (two-stage least squares estimates) and 8 (reduced form estimates).

The results for the social interaction coefficient are smaller than in the previous specifications, suggesting that part of the estimated effect was in fact due to unobserved changes in labor conditions in the local geographic areas. The estimate based on zip codes with at least 100 individuals corresponds to 0.606 ($t-stat = 3.11$) in the model with fixed effects and to 0.651 ($t-stat = 4.07$) in the previous model.

As before, the social interaction coefficient from the sample with at least 50 individuals per zip code is significantly smaller than the baseline sample (with at least 100 individuals per zip code) and is not statistically significant at the 5% confidence level. The implied Social, Private and Spillover effects of the program for this sample (two-stage least squares, more than 50 individuals

[21] The standard errors for the structural parameter β were estimated by delta method.

per zip code) are the following (standard errors in parenthesis):

$$\begin{aligned}
\text{Overall social Impact of the EITC} &= 0.032 * 0.691 = 0.022 \ \ (0.013) \\
\text{Overall private effect of the EITC} &= 0.016 * 0.691 = 0.011 \ \ (0.002) \\
\text{Overall spillover effect of the EITC} &= 0.022 - 0.011 = 0.011 \ \ (0.013)
\end{aligned}$$

As with previous specifications, the R-squared associated with the excluded regressors are small (0.004 for the case with more than 50 individuals per zip code), suggesting a potentially weak instrument problem. Assuming that the instrument is a valid one (uncorrelated with the main error term), we have to be careful with the statistical inference based on the t-statistic, which might be biased in small samples.

5 Conclusions

In this article, we analyze the social impact of the EITC, which includes both the direct effect that has been repeatedly found in earlier evaluations of the EITC, and the social interaction or spillover effect that results when some women start leaving welfare for work, leading other women to follow their steps. Our direct results are extremely robust and similar to the earlier literature but our social interaction estimates suggest that the indirect effects (multiplier) could double the direct impact, implying that the EITC expansion could be responsible for an overall increase between 2.2 and 3.1 percentage points in the labor force participation of low income single mothers in California. The spillover effects are less precisely estimated, however, being in some cases not significantly different from zero.

An open question in this, as in other social interaction studies, is related to the channel through which these social interactions are operating. The EITC is a very expensive program and it is not clear whether expanding it (for instance through a State EITC) would produce similar results on labor force participation to when the program was initially implemented. If we were able to find that diffusion of information plays an important role, it would be possible to design more cost-effective policies that would operate directly on the social interactions (like job search services in low participation areas). In the research agenda on this topic, we will study whether the existence of tax preparation services (TPS), one of the main avenues through which low income individuals claim the EITC, has a role as a facilitator of information transmission; the easiest way to convey information about how to apply for the EITC is to refer other people to the closest TPS in the area.

Another implication of these findings is related to the topic of evaluation of government programs. Quite possibly, few things influence more the policy discussion on government programs than experimental evaluations. Given the complex selection issues involved in voluntary participation, a policy experiment has many desirable properties. However, a critical and often overlooked assumption behind social experiments is the non-existence of social interaction effects.[22] Given the magnitude of the spillover effects found in this paper, it is important to incorporate in the experimental protocol, features that allow measurement of community level externalities.[23]

This paper contributes to empirical literature on social interactions, using an approach based on the assumption of a linear model. Departures from this assumption will likely generate multiple equilibria. Identification of such models is still an open question in the econometric literature. Further research in this area will allow us to further understand the channels through which the environment of an individual can affect the choices she makes.

6 Appendix A: Overview of the EITC[24]

The Earned Income Tax Credit is a Federal Program created in 1975, that provides a subsidy for low-income families with children. Initially small, the program was greatly expanded in the early 1990s to become the largest cash transfer program in the US.

It is operated through the income tax system; a family (either a married couple or a head of household) with children that had some work income during a year can file, even if they do not owe any taxes, an income tax form claiming the credit. The benefit works as a refundable credit: if the amount of the benefit is higher than the tax liability, a check is issued for the difference.

In Figure 1, we show the benefit schedule of the program. Starting at zero earnings, the level of benefits increases with earnings up to a certain point. This is called the "phase-in range". Benefits remain constant during the "flat range" and are then reduced during the "phase-out" range.

In contrast with traditional social programs, the EITC provides an unambiguous incentive to participate in the labor market, at least for single headed households. This can be seen in Figure 2, where we show the traditional consumption-leisure trade-off model, comparing the regular budget constraint (without EITC) and the modified budget constraint.

Two additional characteristics should be mentioned about the program that separate it from traditional welfare programs: the program does not distinguish between single parents or married couples and the benefits are capped for family size, i.e. the amount is the same for families with 2 or more children.

[22] See for example Heckman and Smith (1995) or Burtles (1995) for two discussions about the validity of experimental evaluation of social programs.

[23] A famous example of this approach is the Mexican PROGRESSA program to promote education and nutrition in rural households, in which communities were randomized before the implementation of the program (see Skoufias (2001) for more details on the program and results). Another example is Duflo and Saez (2003) were different departments in a University were randomly selected as well as individuals within the chosen Departments.

[24] See Hotz and Schoolz (2000) for an extensive review of the EITC, its history and behavioral effects.

7 Appendix B: Identification of Social Interaction effects with Panel Data

In this appendix, we will make a first attempt to establish the conditions under which a fixed effect approach at the individual level can arguably reduce or eliminate the endogeneity bias caused by endogenous membership[25].

We will first discuss the simple case with exogenous membership, after which we will analyze the more realistic case of endogenous self-selection into neighborhoods.

7.1 Exogenous membership and exogenous observed characteristics

The simplest model of social interaction would assume that individuals are exogenously assigned into neighborhoods and that the observed determinants (X) are uncorrelated with the unobserved determinants. By exogenous neighborhoods, we mean that the location process is independent of the observed (X) or unobserved determinants (ε) of the outcome of interest and the unobserved neighborhood characteristics (μ).

$$\begin{aligned}
&Y_{igt} = \alpha + \beta\overline{Y_{gt}} + \gamma X_{igt} + \left(\mu_{gt} + \varepsilon_{igt}\right) \qquad (7)\\
&E\left[\mu_{gt} | \left\{X_{igt}\right\}_{i=1}^{N_{gt}}, G_{it} = g\right] = 0\\
&E\left[\varepsilon_{igt} | \left\{X_{igt}\right\}_{i=1}^{N_{gt}}, G_{it} = g\right] = 0\\
&G_{it} \perp \varepsilon_{igt}, \mu_{gt}, X_{igt}
\end{aligned}$$

Where Y_{igt} is the outcome of interest, $\overline{Y_{gt}}$ is the contemporaneous mean of the outcome variable, X_{igt} is an individual determinant of the decision, ε_{igt} is an error term, and μ_{gt} represents unobserved group components. β represents the (endogenous) social interaction effect. G_{it} represents the location of individual i in time t. N_g represents the number of individuals in neighborhood g at time t.

Even under this situation, equation (7) cannot be consistently estimated by ols, because of the simultaneity of the system of equations for individuals belonging to the same group and because the group level mean is correlated with the unobserved group component, as we will see in the following equation.

By inverting the linear system of equations we can solve for the social interaction equilibrium level, which constitutes the first reduced form equation:

$$RF_1 : \overline{Y_{gt}} = \frac{\alpha}{1-\beta} + \frac{\gamma}{1-\beta}\overline{X_{gt}} + \left(\frac{1}{1-\beta}\mu_{gt} + \frac{1}{1-\beta}\overline{\varepsilon_{gt}}\right) \qquad (8)$$

Notice that the parameters in this equation can be consistently estimated given that the error terms have conditional mean equal to zero. However, we cannot separately identify the social

[25] It is also possible that, under more stringent conditions, a repeated cross-section approach might be able to identify the social interaction effects. We will not discuss that possibility in this article.

interaction effect from this equation alone.

$$
\begin{aligned}
E\left[\mu_{gt} \middle| \{X_{igt}\}_{i=1}^{N_{gt}}, G_{it}=g\right] = 0 \quad &\Rightarrow \quad E\left[\mu_{gt}|\overline{X_{gt}}\right] = 0 \\
E\left[\varepsilon_{igt} \middle| \{X_{igt}\}_{i=1}^{N_{gt}}, G_{it}=g\right] = 0 \quad &\Rightarrow \quad E\left[\overline{\varepsilon_{gt}}|\overline{X_{gt}}\right] = 0 \\
\overline{Y_{gt}} = \pi_0 + \pi_1 \overline{X_{gt}} + \nu_{gt} \quad &\Rightarrow \quad \begin{cases} \pi_0 = \frac{\alpha}{1-\beta} \\ \pi_1 = \frac{\gamma}{1-\beta} \end{cases}
\end{aligned} \tag{9}
$$

By replacing the equilibrium equation in (7), we obtain the second reduced form equation.

$$
RF_2 : Y_{igt} = \frac{\alpha}{1-\beta} + \frac{\beta\gamma}{1-\beta}\overline{X_{gt}} + \gamma X_{igt} + \left(\frac{\beta}{1-\beta}\mu_{gt} + \varepsilon_{igt} + \frac{\beta}{1-\beta}\overline{\varepsilon_{gt}}\right) \tag{10}
$$

This time, the exogeneity assumptions will allow us to consistently estimate the coefficients in equation (10) by ols. Notice that the variance covariance matrix will have a block diagonal structure, given that the error terms will be correlated across individuals in the same neighborhood, due to the presence of μ_{gt} and $\overline{\varepsilon_{gt}}$.[26]

$$
\begin{aligned}
E\left[\mu_{gt} \middle| \{X_{igt}\}_{i=1}^{N_{gt}}, G_{it}=g\right] = 0 \quad &\Rightarrow \quad E\left[\mu_{gt}|\overline{X_{gt}}, X_{igt}, G_{it}=g\right] = 0 \\
E\left[\varepsilon_{igt} \middle| \{X_{igt}\}_{i=1}^{N_{gt}}, G_{it}=g\right] = 0 \quad &\Rightarrow \quad \begin{cases} E\left[\varepsilon_{igt}|\overline{X_{gt}}, X_{igt}, G_{it}=g\right] = 0 \\ E\left[\overline{\varepsilon_{gt}}|\overline{X_{gt}}, X_{igt}, G_{it}=g\right] = 0 \end{cases} \\
Y_{igt} = \pi_2 + \pi_3 \overline{X_{gt}} + \pi_4 X_{igt} + \omega_{igt} \quad &\Rightarrow \quad \begin{cases} \pi_2 = \frac{\alpha}{1-\beta} \\ \pi_3 = \frac{\beta\gamma}{1-\beta} \\ \pi_4 = \gamma \end{cases}
\end{aligned} \tag{11}
$$

Finally, a minimum distance estimator could be used to efficiently estimate the coefficients of interest from conditions (9) and (11).

An alternative to this method would be to estimate equation (8) by ols and use the predicted value $\widehat{\overline{Y_{gt}}} = \hat{\pi}_0 + \hat{\pi}_1 \overline{X_{gt}}$ instead of $\overline{Y_{gt}}$ in equation (7). By construction, $\widehat{\overline{Y_{gt}}}$ will be orthogonal to μ_{gt} and this two-stage least squares procedures will provide another consistent estimator of the social interaction effect.

7.2 Endogenous membership and individual fixed effects

If assignment into neighborhoods is non-random, i.e. if individuals self-select into neighborhoods on the basis of characteristics of the neighborhoods (cost of living, amenities, characteristics of current residents) and individual preferences for the different characteristics, then the error term in the structural equation could potentially have a different distribution across neighborhoods. If this error term is correlated with the observed individual characteristics (X_{igt}), or with the

[26] Furthermore, assuming homoskedasticity in μ_{gt} and ε_{igt} the covariance structure will have a known form so that the equation can, in principle, be efficiently estimated by feasible generalized least squares (in a first stage β, σ^2_ε and σ^2_μ can be consistently estimated by ols).

aggregate measure ($\overline{X_{gt}}$), the estimator discussed in the previous section would not be consistent. We will suggest a general model of neighborhood location in which the individuals are maximizing a neighborhood specific utility, which is a function of neighborhood characteristics and individual valuations ($\delta' W_{igt}$), an individual*neighborhood effect (ϕ_{ig}) and a neighborhood specific utility shock (η_{igt}). Some components in the utility function (η and ϕ) will be correlated with the error term in the main equation (ε), giving rise to the selection problem.

$$
\begin{aligned}
& G_{it} = \arg \max_{g \in \{A,B,\ldots\}} \left\{ U_{igt} = \delta' W_{igt} + \phi_{ig} + \eta_{igt} \right\} \\
& Y_{igt} = \alpha + \beta \overline{Y_{gt}} + \gamma X_{igt} + \left(\psi_{ig} + \mu_{gt} + \varepsilon_{igt} \right) \qquad (12) \\
& \left(\varepsilon_{igt}, \left\{ \eta_{igt} \right\}_{g \in \{A,B,\ldots\}} \right) \sim i.i.d. \\
& E\left[\mu_{gt} | \left\{ X_{igt} \right\}_{i=1}^{N_{gt}}, \left\{ \phi_{ig}, \psi_{ig} \right\}_{g \in \{A,B,\ldots\}}, G_{it} = g \right] \neq 0 \\
& E\left[\varepsilon_{igt} | \left\{ X_{igt} \right\}_{i=1}^{N_{gt}}, \left\{ \phi_{ig}, \psi_{ig} \right\}_{g \in \{A,B,\ldots\}}, G_{it} = g \right] \neq 0 \\
& E\left[X_{igt} \psi_{ig} \right] \neq 0 \qquad (13) \\
& \eta_{igt} \not\perp \varepsilon_{igt}, \mu_{gt}, X_{igt}
\end{aligned}
$$

Notice that we allowed for individual*neighborhood specific fixed effects in equation (12).

We are interested in understanding the conditions under which the use of individual fixed effects could eliminate the selection problem in this framework. Assuming that we have access to a two-period panel of individuals, we can express the fixed effect model as a first difference model. Equation (12) *for those individuals that didn't change neighborhood* between periods 1 and 2 corresponds to:

$$
\begin{aligned}
\text{If} \qquad & G_{i1} = G_{i2} = g \\
& \Delta Y_{ig} = \beta \Delta \overline{Y_g} + \gamma \Delta X_{ig} + \left(\Delta \mu_g + \Delta \varepsilon_{ig} \right) \qquad (14) \\
\Rightarrow \quad & \begin{cases} \Delta \overline{Y_g} = \frac{\gamma}{1-\beta} \Delta \overline{X_g} + \left(\frac{1}{1-\beta} \Delta \mu_g + \frac{1}{1-\beta} \Delta \overline{\varepsilon_g} \right) \\ \Delta Y_{ig} = \frac{\beta\gamma}{1-\beta} \Delta \overline{X_g} + \gamma \Delta X_{ig} + \left(\frac{\beta}{1-\beta} \Delta \mu_g + \Delta \varepsilon_{ig} + \frac{\beta}{1-\beta} \Delta \overline{\varepsilon_g} \right) \end{cases} \qquad (15) \\
\text{with} \qquad & \Delta Y_{ig} \equiv Y_{ig2} - Y_{ig1} \\
& \Delta \overline{X_g} \equiv \overline{X_{g2}} - \overline{X_{g1}} \\
& \Delta X_{ig} \equiv X_{ig2} - X_{ig1} \\
& \Delta \mu_g \equiv \mu_{g2} - \mu_{g1} \\
& \Delta \varepsilon_{ig} \equiv \varepsilon_{ig2} - \varepsilon_{ig1} \\
& \Delta \overline{\varepsilon_g} \equiv \overline{\varepsilon_{g2}} - \overline{\varepsilon_{g1}}
\end{aligned}
$$

The first difference eliminates the constant term and any time invariant unobserved individual and group characteristics. The first assumption we will introduce is that changes in unobserved group components are orthogonal to the changes in the average observed characteristic:

$$A_1 : E\left[\Delta\mu_g | \Delta\overline{X_g}, \Delta X_{ig}, G_{i1} = G_{i2} = g\right] = 0 \tag{16}$$

The second critical assumption is that the first difference error term has zero mean, conditional on $\Delta\overline{X_g}$, ΔX_{ig} and the location decisions on both periods:

$$A_2 : E\left[\Delta\varepsilon_{ig} | \Delta\overline{X_g}, \Delta X_{ig}, G_{i1} = G_{i2} = g\right] = 0 \tag{17}$$

To see the validity of (A_2), let's introduce the conditions implied by the location decision for the case in which there are only two neighborhoods (A and B):

$$\begin{aligned}
&E\left[\Delta\varepsilon_{ig} | \Delta\overline{X_g}, \Delta X_{ig}, G_{i1} = G_{i2} = A\right] \\
&= E\left[\varepsilon_{ig2} | \Delta\overline{X_g}, \Delta X_{ig}, G_{i2} = A\right] - E\left[\varepsilon_{ig1} | \Delta\overline{X_g}, \Delta X_{ig}, G_{i1} = A\right] \\
&= E\left[\varepsilon_{ig2} | \Delta\overline{X_g}, \Delta X_{ig}, \delta' W_{iA2} + \eta_{iA2} > \delta' W_{iB2} + \eta_{iB2}\right] \\
&\qquad - E\left[\varepsilon_{ig1} | \Delta\overline{X_g}, \Delta X_{ig}, \delta' W_{iA1} + \eta_{iA1} > \delta' W_{iB1} + \eta_{iB1}\right] \\
&= E\left[\varepsilon_{ig2} | \Delta\overline{X_g}, \Delta X_{ig}, \delta'\left(W_{iA2} - W_{iB2}\right) + \left(\eta_{iA2} - \eta_{iB2}\right) > 0\right] \\
&\qquad - E\left[\varepsilon_{ig1} | \Delta\overline{X_g}, \Delta X_{ig}, \delta'\left(W_{iA1} - W_{iB1}\right) + \left(\eta_{iA1} - \eta_{iB1}\right) > 0\right] \\
&= E\left[\varepsilon_{ig2} | \Delta\overline{X_g}, \Delta X_{ig}, \delta'\tilde{W}_{i2} + \tilde{\eta}_{i2} > 0\right] - E\left[\varepsilon_{ig1} | \Delta\overline{X_g}, \Delta X_{ig}, \delta'\tilde{W}_{i1} + \tilde{\eta}_{i1} > 0\right] \\
&= 0 \text{ if } \delta'\tilde{W}_{i2} = \delta'\tilde{W}_{i1} \text{ and the distribution of } \varepsilon \text{ remains stationary between } t = 1 \text{ and } t = 2
\end{aligned}$$

In other words, it is required that the interaction of neighborhood characteristics and individual preferences remains reasonably constant over the time periods under consideration. It is also required that the changes in the average observed characteristics are not accompanied by changes in the distribution of unobserved characteristics.

Under assumptions A_1 and A_2, we can estimate the reduced form equations in (15) and obtain consistent estimates for the parameters of interest, using the same ideas for the model with exogenous membership.

One potential problem with the previous reasoning is that some of the underlying differences across neighborhoods might actually change over time, particularly job opportunities. These differences will cause movements of individuals that might be correlated with changes in the average program or labor force participation levels. For that reason, it is important to include controls for the local economic conditions on the relevant labor markets. Furthermore, using the Instrumental Variable approach introduced in the previous section should eliminate most of the remaining endogeneity, as long as the decisions to stay or move are not systematically related to the instrument.

One might argue that the previous model is very specific in the structure of the location endogeneity. However, the basic idea is that if the location decisions of the individuals are not drastically altered (except for idiosyncratic shocks) by the intervention that occurs between the two periods, the selectivity bias term can be 'differenced out'. More generally, we require that the speed of adjustment of the location process be slower than the social interaction effect[27].

To better understand the argument, let's suppose that after the exogenous intervention occurs, the social interaction process alters the participation decisions of individuals and as a result a significant fraction of the population modify their location decision *during* t=2. The new allocation will reflect a new equilibrium of the location process (different to the one in t=1) and the new mix of individuals might have different properties in terms of the unobserved determinants of participation. Taking first differences would not eliminate the selectivity bias in this case.

Let's not forget that, even though this discussion suggests that the fixed effect approach might help eliminate the endogeneity bias, it will potentially eliminate the variation in the average participation component (or more specifically in the average value of the exogenous variable X, which acts as instrument for mean participation). For that reason, it is still important to find an intervention that differentially affects the different neighborhoods contemporaneously and over time.

[27] The role of speed of adjustment has been reported in the identification literature but never been explicitly modeled.

8 Appendix C: A simple model of neighborhood location and labor force participation

We present a simple model that combines neighborhood choice with the decision to participate in the labor market. The idea of the model is to illustrate the fact that endogenous group formation can often lead to correlation in unobservable (and observable) characteristics among individuals in the same group. As we will see, this correlation will generate a relationship between individual and group behavior, without any social interaction in the model.

In a city consisting of 2 neighborhoods and 2 possible states (work or not work), each woman has to make 2 decisions in each period:

1. Whether she will work for a fixed wage W or not work and collect welfare cash benefits for an amount of B (with $B < W$).

2. Whether to live in neighborhood A or B. Depending on what neighborhood she chooses, she will have to pay rent R_A or R_B.

The individual takes these parameters as given (W, B , R_A or R_B).

These decisions will be made in order to maximize a utility function, which depends on the amount of consumption the person can afford with the income net of rent payments, the leisure time that remains from the work decision, and utility from a neighborhood specific attribute, in this case, air quality. We will assume that neighborhood B has good air quality and neighborhood A doesn't.

$$\begin{aligned} U_{itn} &= U\left(C_{itn}, L_{itn}, Q_{tn}\right) \\ C_{itn} &= \begin{cases} W - R_n & \text{if work } (Y_{it} = 1) \\ B - R_n & \text{if not } (Y_{it} = 0) \end{cases} \\ L_{itn} &= \begin{cases} 0 & \text{if work } (Y_{it} = 1) \\ 1 & \text{if not } (Y_{it} = 0) \end{cases} \\ Q_{itn} &= \begin{cases} 0 & \text{if live in neighborhood A } (A_{it} = 1) \\ 1 & \text{if live in neighborhood B } (A_{it} = 0) \end{cases} \end{aligned}$$

To make the model tractable, we will assume that the utility function is additive in the different terms, according to the following specification:

$$U_{itn} = \alpha_i C_{itn} + (1 - \alpha_i) L_{itn} + \rho_i Q_{itn}$$

Where ρ_i and α_i are individual specific preference parameters independent of each other, uniformly distributed in the $[0, 1]$ interval and assumed constant in the short run.

A full general equilibrium model would require me to specify how the wage level (W), the welfare benefit level (B) and the rent differential between the neighborhoods ($R_B - R_A$) are set.

For this discussion, we will assume that the system is in equilibrium and that W, B and $(R_B - R_A)$ are given.

Given that neighborhood B has a special feature that is valued positively by all individuals, we can safely assume that in equilibrium, R_B will be higher than R_A.

Given the linearity of the problem, the two decisions can be analyzed separately. This can clearly be seen in the following table:

Utilities perceived from the 4 possible choices

	Neighborhood A	Neighborhood B
Work	$\alpha\left(W - R_A\right)$	$\alpha\left(W - R_B\right) + \rho$
Not work	$\alpha\left(B - R_A\right) + (1 - \alpha)$	$\alpha\left(B - R_B\right) + \rho + (1 - \alpha)$

Independently of the neighborhood where the individual ultimately choose to live, the work decision is given by the following condition which states that only individuals with a high preference for work relative to leisure (α) will choose to do so:

$$Y_{itn} = 1 \Leftrightarrow \alpha \geq \frac{1}{W - B + 1}$$

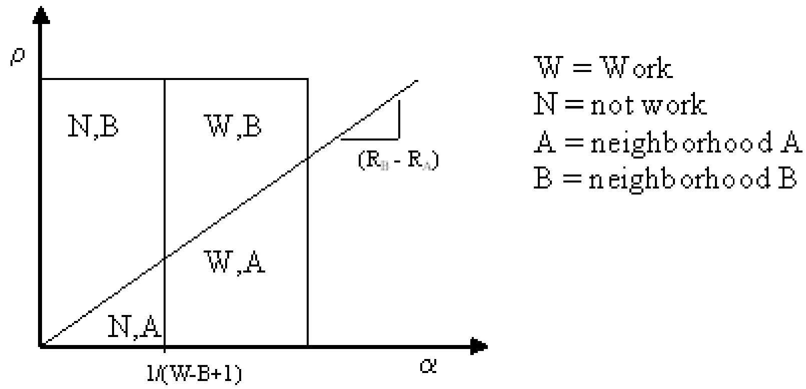
Similarly, independently of the work decision, the location decision is given by the following condition, which implies that only individuals with a high preference for Air Quality (ρ) will be willing to pay the extra rent cost associated with living in neighborhood B.

$$A_{it} = 1 \Leftrightarrow \alpha \geq \frac{\rho}{(R_B - R_A)}$$

An important feature of this equation is that the (unobserved) marginal utility of consumption, α, is also part of the location decision, which means that the average level of α will be higher in neighborhood A than in neighborhood B.

The previous solution can be expressed in the following graph, showing what type of individuals would work and choose to live in the more expensive neighborhood.

The sorting equilibrium into neighborhoods and jobs (or welfare)



Given the uniformity and independence assumptions for α and ρ, the average labor force participation in each neighborhood are given by the ratios of the following areas in the previous graph:

$$
\begin{aligned}
\overline{Y_A} &= \frac{Area\,(W,A)}{Area\,(W,A)+Area\,(N,A)} \\
\overline{Y_B} &= \frac{Area\,(W,B)}{Area\,(W,B)+Area\,(N,B)}
\end{aligned}
$$

From the previous graph we can appreciate that $\overline{Y_A} \geq \overline{Y_B}$. If an individual has a high preference for consumption relative to leisure (α), she is more likely to be living in neighborhood A, where average participation is higher.

This suggests that, in a model with more neighborhoods, a simple regression estimation of individual participation as a function of $\bar{Y}_n$ (the average labor force participation in the neighborhood chosen by the individual) and other determinants of work will return a positive coefficient on $\bar{Y}_n$, *without there being any social interaction present.*

References

[1] Aaronson, D., "Sibling estimates of neighborhood effects", in Brooks-Gunn, J., G.J. Duncan, & L. Aber, Eds. Neighborhood Poverty: Policy Implications in Studying Neighborhoods, Vol II. (pp. 80-93). New York: Russell Sage Foundation.

[2] Aizer, A. and Currie, J., "Networks or Neighborhoods? Correlations in the Use of Publicly-Funded Maternity Care in California", NBER Working Paper No.w9209, September 2002.

[3] Bertrand, Luttmer and Mullainathan, "Network Effects and Welfare Cultures", Quarterly Journal of Economics 115, August 2000, pages 1019-1056.

[4] Brock, William A. and Durlauf, Steven N., "Interactions-Based Models", Handbook of Econometrics, Volume 5, Edited by J.J. Heckman and E. Leamer, 2001.

[5] Burtless, Gary, "The Case for Randomized Field Trials in Economics and Policy Research", The Journal of Economic Perspectives, Volume 9, Issue 2 (Spring, 1995), pages 63-84.

[6] Case, A. and L. Katz, "The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youths", NBER Working Paper No. 9375, May 1991.

[7] Cutler, D and E. Glaeser, "Are Ghettos Good or Bad", Quarterly Journal of Economics 112, August 1997, pages 827-872.

[8] Dickert, Stacy, Scott Houser and John Karl Scholz, 1995, "The Earned Income Tax Credit and Transfer Programs: A Study of Labor Market and Program Participation, Tax Policy and the Economy", James M. Poterba (ed.), National Bureau of Economic Research and the MIT Press 9, 1-50.

[9] Duflo, E. and Saez, E., "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment", NBER Working Paper No.w8885, April 2002.

[10] Duncan, Greg and Raudenbush, Stephen, "Neighborhoods and Adolescent Development: How Can we Determine the Links?" working paper, Northwestern University.

[11] Eissa, N. and Hoynes, H., "The Earned Income Tax Credit and the Labor Supply of Married Couples", National Bureau of Economic Research Working Paper: 6856, December 1998.

[12] Eissa, Nada and Jeffrey B. Liebman, "Labor Supply Response to the Earned Income Tax Credit", Quarterly Journal of Economics, May 1996, 605-637.

[13] Ellwood, David T., 2000, "The Impact of the Earned Income Tax Credit and Social Policy Reforms on Work, Marriage, and Living Arraignments", mimeo, Kennedy School of Government.

[14] Evans, Oates and Schwab, “Measuring Peer Group Effects: A Study of Teenage Behavior”, Journal of Political Economy, 1992, vol. 100, no. 5.

[15] Glaeser, Sacerdote and Scheinkman, “Crime and Social Interactions”, Quarterly Journal of Economics, May 1996, pages 507-548.

[16] Glaeser, Sacerdote and Scheinkman, “The Social Multiplier”, NBER Working Paper No.w9153, September 2002.

[17] Grogger, Jeffrey, "Welfare Transitions in the 1990s: The Economy, Welfare Policy, and the EITC", NBER Working Paper w9472, February 2003.

[18] Grogger, Jeffrey, "The Effects of Time Limits, the EITC, and Other Policy Changes on Welfare Use, Work, and Income among Female-Headed Families", The Review of Economics and Statistics, Vol. 85, Issue 2, May 2003.

[19] Heckman James and Smith, Jeffrey, “Assessing the Case for Social Experiments”, The Journal of Economic Perspectives, Volume 9, Issue 2 (Spring, 1995), pages 85-110.

[20] Hotz, Joseph V. and Scholz, John Karl, “The Earned Income Tax Credit”, forthcoming in Means-Tested Transfer Programs in the United States.

[21] Hotz, V. Joseph, Charles H. Mullin and John Karl Scholz, “The Earned Income Tax Credit and Labor Market Participation of Families on Welfare”, mimeo, UCLA, Vanderbilt and University of Wisconsin Madison, March 2002.

[22] Katz, Lawrence, Kling, Jeffrey and Liebman, Jeffrey, “Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment”, NBER Working paper 7973, October 2000.

[23] Kooreman, P. and Soetevent, A., “A Discrete Choice Model with Social Interactions; an Analysis of High School Teen Behavior”, mimeo, University of Groningen, February 2002.

[24] Kyriazidou, Ekaterini, “Estimation of A Panel Data Sample Selection Model,” Econometrica, Vol. 65, No 6, pp. 1335-1364 (1997).

[25] Ludwig, Jens, Duncan, Greg, and Pinkston, Joshua, “Neighborhood Effects on Economic Self-Sufficiency: Evidence from a Randomized Housing-Mobility Experiment”, mimeo Georgetown Public Policy Institute, January 31, 2000.

[26] Charles Manski, “Identification of Endogenous Social Effects: The Reflection Problem”, Review of Economic Studies, 60, 1993, pages 531-542.

[27] Charles Manski, “Economic Analysis of Social Interactions”, Journal of Economic Perspectives, forthcoming.

[28] Marmaros, David and Sacerdote, Bruce, “Peer and social networks in job search”, European Economic Review 46 (2002), 870-879.

[29] Meyer, B. and Rosenbaum, D., "Welfare, the Earned Income Tax Credit, and the Labor Supply of Single Mothers", Quarterly Journal of Economics, August 2001; 116(3): 1063-1114.

[30] Robert Moffitt, "Policy Interventions, Low-Level Equilibria and Social Interactions", Forthcoming In Social Dynamics, eds. Steven Durlauf and Peyton Young, MIT Press.

[31] Moretti, E., "Estimating the Social Return to Higher Education: Evidence From Longitudinal and Cross-Sectional Data", forthcoming in Journal of Econometrics.

[32] Sacerdote, B., "Peer Effects with Random Assignment: Results for Dartmouth Roommates", Quarterly Journal of Economics. May 2001; 116(2): 681-704.

Figure 1 – EITC Benefit schedule

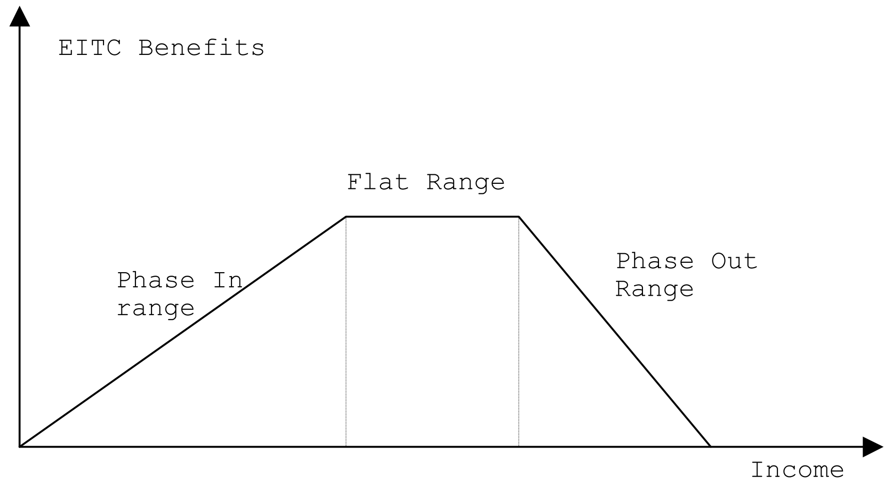


Figure 2 – The work incentive of the EITC

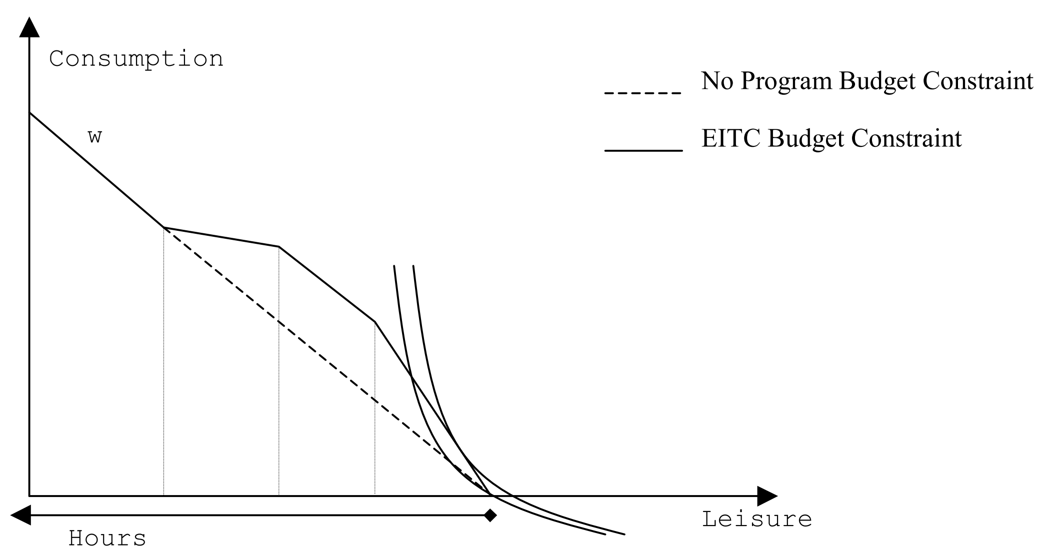


Figure 3. Distribution of average change in LFP, zip code, 93-94

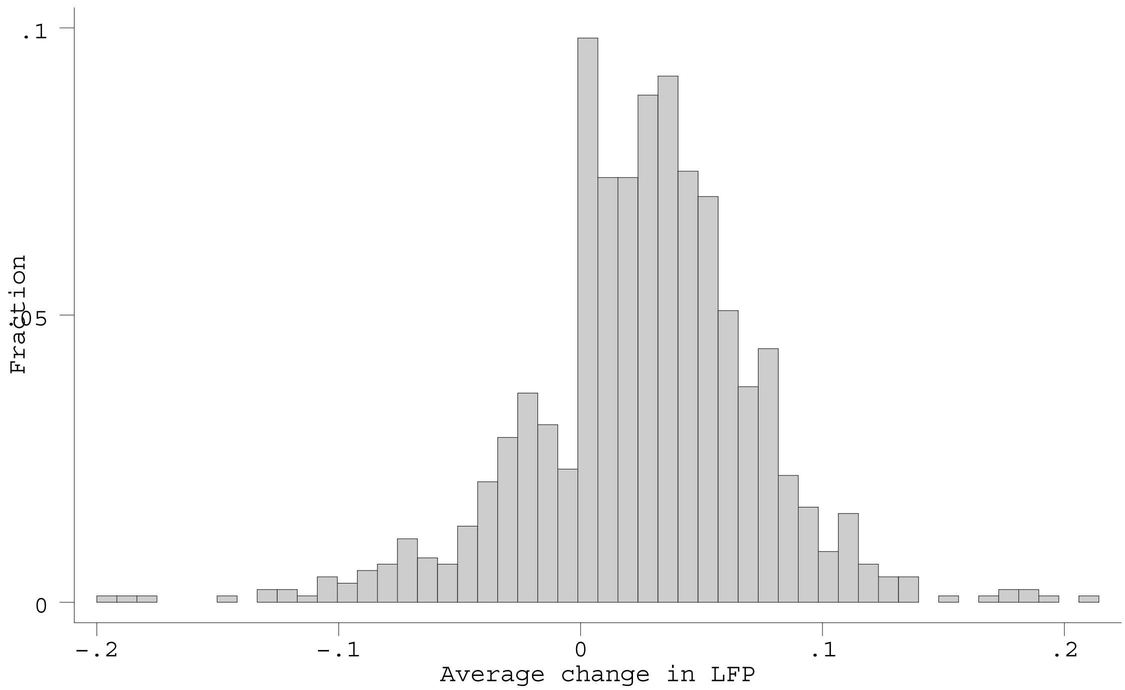


Figure 4. Distribution of average fraction 1 vs 2+ chidren, 93-94

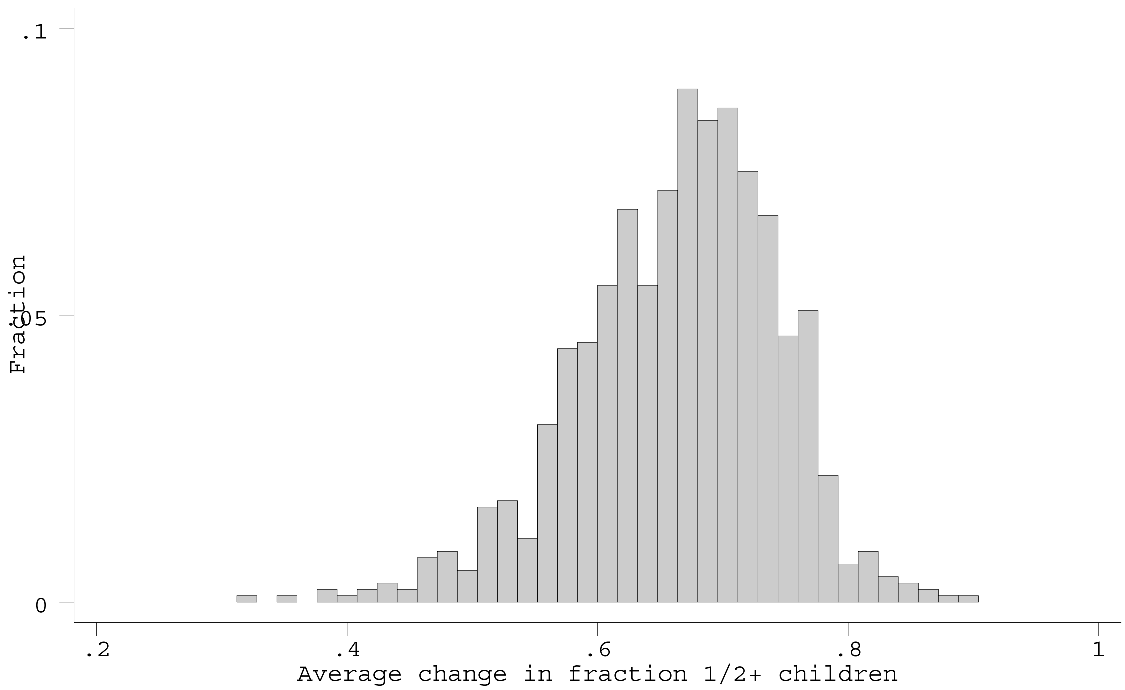


Table 1 – EITC parameters between 1993 and 1995

	Phase-In Region	Flat Region			Phase-Out Region	
Year	Credit Rate	Begining	Ending	Maximum Credit	Phase Out Credit rate	Maximum Income
1993						
1 child	18.5%	$7,750	$12,200	$1,434	13.21%	$23,050
2+ children	19.5%	$7,750	$12,200	$1,511	13.93%	$23,050
1994						
1 child	26.30%	$7,750	$11,000	$2,038	15.98%	$23,755
2+ children	30.0%	$8,425	$11,000	$2,528	17.68%	$25,296
No children	7.65%	$4,000	$5,000	$306	7.65%	$9,000
1995						
1 child	34.0%	$6,160	$11,290	$2,094	15.98%	$24,396
2+ children	36.0%	$8,640	$11,290	$3,110	20.22%	$26,673
No children	7.65%	$4,100	$5,130	$314	7.65%	$9,230

Table 2 - Summary statistics - baseline sample

	1993		1994	
	mean	s.e.	mean	s.e.
Families				
Number of observations	142391		142391	
Labor Force Participation	0.337	0.473	0.367	0.482
Number of children	2.220	1.120	2.186	1.091
More than 1 child	0.691	0.462	0.691	0.462
Mother of white race	0.288	0.453		
Mother of black race	0.195	0.396		
Mother of Hispanic origin	0.428	0.495		
Mother's age in 1994	33.89	5.95		
County Employment / Population, all sectors	0.603	0.091	0.610	0.091
County Employment / Population, retail sector	0.104	0.010	0.104	0.010
County Employment / Population, services sector	0.161	0.046	0.164	0.046
Unemployment rate	10.27	3.32	9.44	3.04
Zip codes				
Number of zip codes	503		503	
Average number of individuals	283.1	176.7	283.1	176.7
Minimum number of individuals	101		101	
Maximum number of individuals	1255		1255	
Average LFP	0.351	0.076	0.381	0.075
Minimum LFP	0.138		0.161	
Maximum LFP	0.580		0.580	
Average Fraction 1 vs 2+ children families	0.679	0.060	0.679	0.060
Minimum Fraction 1 vs 2+ children families	0.386		0.386	
MaximumFraction 1 vs 2+ children families	0.838		0.838	

Table 3 - Estimate of the Private Effect of the EITC (without Social Interactions)

Dependent Variable = Labor Force Participation (Annual Earnings > $200)

	Cross Section	Pooled 1993 and 1994	Pooled 1993/94 + FE
(More than 1 child)*(Post 93)		**0.016**	**0.016**
(Private effect of EITC)		**(7.11)****	**(7.07)****
More than 1 child	-0.027	-0.043	
	(9.51)**	(15.51)**	
Post 93		0.019	
		(9.85)**	
Mother's age	0.015	0.017	
	(7.64)**	(9.45)**	
Mother's squared age	-0.000	-0.000	-0.000
	(7.22)**	(8.84)**	(2.27)*
Employment / Population - all sectors	-0.072	-0.095	-0.101
	(5.16)**	(7.44)**	(1.78)
Observations	142391	284782	142391
R-squared	0.0012	0.0029	0.0004

Absolute value of t statistics in parentheses (standard errors in pooled regression clustered by individual)
* significant at 5%; ** significant at 1%

Table 4 - Estimate of the Private, Spillover and Social Effects of the EITC

Dependent Variable = Labor Force Participation (Annual Earnings > $200)

	FE without IV	IV without FE	Baseline: IV + FE
Average LFP	**0.996**	**0.753**	**0.651**
(Spillover effect)	**(472.79)****	**(8.47)****	**(4.07)****
(More than 1 child)*(Post 93)	**0.015**	**0.015**	**0.016**
(Private effect of EITC)	**(6.59)****	**(3.83)****	**(6.59)****
More than 1 child		-0.041	
		(14.24)**	
Post 1993		-0.004	
		(1.09)	
Mother's age		0.015	
		(9.38)**	
Mother's squared age	-0.000	-0.000	-0.000
	(2.20)*	(8.70)**	(2.23)*
Employment / Population - all sectors	-0.104	-0.026	-0.103
	(1.72)	(2.67)**	(1.70)
# Observations	142391	284782	142391
# zip codes	503	503	503
R-squared	0.0046	0.0249	0.0041

First Stage: Dependent Variable = Average Labor Force Participation in the zip code

Average ((More than 1 child)*(Post 93))	**0.046**
(Social impact of the EITC)	**(2.29)***
R-squared	0.0015
R-squared excluded instrument	0.010

Absolute value of t statistics in parentheses (standard errors allowing for clustering by zip code)
* significant at 5%; ** significant at 1%
Note: Other covariates in first stage (not reported) include all individual variables in second stage

Table 5 - Estimate of the Private, Spillover and Social Effects of the EITC - Robustness to sample definition

Second Stage: Dependent Variable = Labor Force Participation (Annual Earnings > $200)

	zip size > 25	**zip size > 50**	**zip size > 100**	**zip size > 100**	**zip size > 100**
	Base sample	**Base sample**	**Base sample**	**Sample B***	**Sample C***
Average LFP	**0.574**	**0.544**	**0.651**	**0.629**	**0.603**
(Spillover effect)	**(3.20)****	**(2.32)***	**(4.07)****	**(1.93)**	**(1.56)**
(More than 1 child)*(Post 93)	**0.017**	**0.016**	**0.016**	**0.012**	**0.012**
(Private effect of EITC)	**(7.70)****	**(7.06)****	**(6.59)****	**(3.55)****	**(3.46)****
Squared age	-0.023	-0.022	-0.019	-0.057	-0.063
	(2.90)**	(2.64)**	(2.23)*	(4.48)**	(4.79)**
Employment / Population - all sectors	-0.111	-0.103	-0.103	-0.091	-0.084
	(2.00)*	(1.80)	(1.70)	(1.33)	(1.21)
# Observations	165139	157762	142391	84853	82146
# zip codes	921	716	503	364	354
R-squared	0.0060	0.0047	0.0041	0.0048	0.0048

First Stage: Dependent Variable = Average Labor Force Participation in the zip code

Average ((More than 1 child)*(Post 93))	**0.041**	**0.036**	**0.046**	**0.034**	**0.031**
(Social impact of the EITC)	**(2.49)***	**(2.01)***	**(2.29)***	**(1.21)**	**(1.08)**
R-squared	0.0066	0.0057	0.0102	0.0052	0.0045
R-squared excluded instrument	0.0060	0.006	0.010	0.004	0.003

Absolute value of t statistics in parentheses (standard errors allowing for clustering by zip code)

* significant at 5%; ** significant at 1%

Note: Other covariates in first stage (not reported) include all individual variables in second stage

Sample B = Base sample restricted to individuals on welfare at least twice between 1992 and 1995

Sample C = Base sample restricted to individuals on welfare at least twice between 1992 and 1995, at most one year apart

Table 6 - Reduced Form estimates, assuming Social Interaction equilibrium

Reduced Form Equation: Dependent Variable = Labor Force Participation (Annual Earnings > $200)

	zip size > 25	zip size > 50	zip size > 100	zip size > 100	zip size > 100
	Base sample	**Base sample**	**Base sample**	**Sample B***	**Sample C***
Average ((More than 1 child)*(Post 93)) (π_1)	**0.02332**	**0.01944**	**0.03020**	**0.02122**	**0.01882**
Implied Spillover effect ($\beta\gamma / (1-\beta)$)	**(1.42)**	**(1.09)**	**(1.48)**	**(0.75)**	**(0.65)**
(More than 1 child)*(Post 93) (π_2)	**0.01671**	**0.01564**	**0.01554**	**0.01196**	**0.01194**
(Private effect of EITC)	**(7.70)****	**(7.07)****	**(6.60)****	**(3.56)****	**(3.47)****
Squared age	-0.00048	-0.00045	-0.00040	-0.00121	-0.00132
	(2.95)**	(2.72)**	(2.28)*	(4.77)**	(5.07)**
Employment / Population - all sectors	-0.10632	-0.09960	-0.09969	-0.06678	-0.06056
	(1.81)	(1.66)	(1.56)	(0.96)	(0.87)
# Observations	165139	157762	142391	84853	82146
# zip codes	921	716	503	364	354
R-squared	0.0005	0.0004	0.0004	0.0005	0.0005

Social interaction estimation, from Reduced Form

Indirect estimate of $\beta=(\pi_1/(\pi_1+\pi_2))$	**0.5826**	**0.5542**	**0.6603**	**0.6395**	**0.6118**
(Social Interaction coefficient)	**(3.28)****	**(2.38)***	**(4.18)****	**(2.00)***	**(1.60)**

Absolute value of t statistics in parentheses (standard errors allowing for clustering by zip code)

* significant at 5%; ** significant at 1%

Sample B = Base sample restricted to individuals on welfare at least twice between 1992 and 1995

Sample C = Base sample restricted to individuals on welfare at least twice between 1992 and 1995, at most one year apart

Table 7 - Models with (county*time) fixed effect - 2SLS estimates

Second Stage: Dependent Variable = Labor Force Participation (Annual Earnings > $200)

	zip size > 25	**zip size > 50**	**zip size > 100**
	Base sample	**Base sample**	**Base sample**
Average LFP	**0.518**	**0.467**	**0.606**
(Spillover effect)	**(2.25)***	**(1.46)**	**(3.11)****
(More than 1 child)*(Post 93)	**0.017**	**0.016**	**0.016**
(Private effect of EITC)	**(7.69)****	**(7.06)****	**(6.60)****
Squared age	-0.000	-0.000	-0.000
	(2.88)**	(2.60)**	(2.21)*
# Observations	165139	157762	142391
# zip codes	921	716	503
R-squared	0.0058	0.0045	0.0040

First Stage: Dependent Variable = Average Labor Force Participation in the zip code

Average ((More than 1 child)*(Post 93))	**0.038**	**0.032**	**0.045**
(Social impact of the EITC)	**(2.19)***	**(1.73)**	**(2.18)***
R-squared	0.0888	0.0962	0.1154
R-squared excluded instrument	0.005	0.004	0.009

Absolute value of t statistics in parentheses (standard errors allowing for clustering by zip code)
* significant at 5%; ** significant at 1%

Table 8 - Models with (county*time) fixed effect - Reduced Form estimates

Reduced Form Equation: Dependent Variable = Labor Force Participation (Annual Earnings > $200)

	zip size > 25	**zip size > 50**	**zip size > 100**
	Base sample	**Base sample**	**Base sample**
Average ((More than 1 child)*(Post 93)) (π_1)	**0.01954**	**0.01513**	**0.02705**
Implied Spillover effect ($\beta\gamma$ / (1−β))	**(1.12)**	**(0.80)**	**(1.30)**
(More than 1 child)*(Post 93) (π_2)	**0.01670**	**0.01564**	**0.01557**
(Private effect of EITC)	**(7.69)****	**(7.06)****	**(6.61)****
Squared age	-0.00047	-0.00044	-0.00039
	(2.92)**	(2.68)**	(2.24)*
# Observations	165139	157762	142391
# zip codes	921	716	503
R-squared	0.0011	0.0010	0.0010

Social interaction estimation, from Reduced Form

Indirect estimate of $\beta=(\pi_1/(\pi_1+\pi_2))$	**0.5392**	**0.4917**	**0.6347**
(Social Interaction coefficient)	**(2.38)***	**(1.54)**	**(3.42)****

Absolute value of t statistics in parentheses (standard errors allowing for clustering by zip code)
* significant at 5%; ** significant at 1%