USING DIFFERENCES IN KNOWLEDGE ACROSS NEIGHBORHOODS TO UNCOVER
THE IMPACTS OF THE EITC ON EARNINGS

Raj Chetty
John N. Friedman
Emmanuel Saez

Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC
on Earnings
Raj Chetty, John N. Friedman, and Emmanuel Saez
NBER Working Paper No. 18232
July 2012
JEL No. H26,H31

ABSTRACT

We develop a new method of estimating the impacts of tax policies that uses areas with little knowledge about the policy's marginal incentives as counterfactuals for behavior in the absence of the policy. We apply this method to characterize the impacts of the Earned Income Tax Credit (EITC) on earnings using administrative tax records covering all EITC-eligible filers from 1996-2009. We begin by developing a proxy for local knowledge about the EITC schedule – the degree of "sharp bunching" at the exact income level that maximizes EITC refunds by individuals who report self-employment income. The degree of self-employed sharp bunching varies significantly across geographical areas in a manner consistent with differences in knowledge. For instance, individuals who move to higher-bunching areas start to report incomes closer to the refund-maximizing level themselves, while those who move to lower-bunching areas do not. Using this proxy for knowledge, we compare W-2 wage earnings distributions across neighborhoods to uncover the impact of the EITC on real earnings. Areas with high self-employed sharp bunching (i.e., high knowledge) exhibit more mass in their W-2 wage earnings distributions around the EITC plateau. Using a quasi-experimental design that accounts for unobservable differences across neighborhoods, we find that changes in EITC incentives triggered by the birth of a child lead to larger wage earnings responses in higher bunching neighborhoods. The increase in EITC refunds comes primarily from intensive-margin increases in earnings in the phase-in region rather than reductions in earnings in the phase-out region. The increase in EITC refunds is commensurate to a phase-in earnings elasticity of 0.14 on average across the U.S. and 0.58 in high-knowledge neighborhoods.

Raj Chetty
Department of Economics
Harvard University
1805 Cambridge St.
Cambridge, MA 02138
and NBER
chetty@fas.harvard.edu

John N. Friedman
Harvard Kennedy School
Taubman 356
79 JFK St.
Cambridge, MA 02138
and NBER
john_friedman@harvard.edu

Emmanuel Saez
Department of Economics
University of California, Berkeley
530 Evans Hall #3880
Berkeley, CA 94720
and NBER
saez@econ.berkeley.edu

I   Introduction

Research on the impacts of tax policies on economic behavior has confronted two important empirical challenges. First, because federal tax policies often do not vary cross-sectionally, it is difficult to find counterfactuals that permit credible estimation of the policies' causal effects (Meyer 1995, Saez et al. 2012). Second, many individuals may respond slowly to tax changes because of inattention to the tax code and other adjustment frictions (Brown 1968, Fujii and Hawley 1988, Bises 1990). This makes it difficult to identify steady-state behavioral responses using short-run comparisons before and after a tax reform (Chetty et al. 2011, Chetty 2012).

We develop a research design that addresses these challenges by exploiting differences across neighborhoods in knowledge about the tax code. Our method is based on a simple idea: individuals with no knowledge of a tax policy's marginal incentives will behave as they would in the absence of the policy.[1] Hence, one can identify the causal effect of a policy by comparing behavior across cities that differ in knowledge about the policy but are otherwise comparable. We apply this method to analyze the impacts of the Earned Income Tax Credit (EITC), the largest means-tested cash transfer program in the United States, on earnings behavior and inequality. We exploit fine geographical heterogeneity across ZIP codes by using selected data from U.S. population tax records spanning 1996-2009, which include over 75 million unique EITC eligible individuals with children and 1 billion observations on their annual earnings. Our method uncovers significant impacts of the EITC on earnings behavior. The intensive-margin responses we document are masked in aggregate data and cannot be easily detected using traditional research designs because of their diffuse nature, potentially explaining why prior studies find mixed evidence of intensive margin responses to the EITC and other tax policies.

Our empirical analysis proceeds in two steps. First, we develop a proxy for local knowledge about the marginal rate structure of the EITC schedule.[2] Ideally, one would measure knowledge directly using data on individuals' perceptions of the EITC schedule. Lacking such data, we proxy for knowledge using the extent to which individuals manipulate their reported income to maximize their EITC refunds by reporting self-employment income. Self-employed tax filers have a propensity

---

[1] As we discuss in Section 2 below, this equivalence holds in the absence of income effects. With income effects, our technique recovers compensated elasticities under the assumption that uninformed individuals believe that the tax credit is a lump-sum subsidy.

[2] Throughout the paper, we use the term "knowledge" or "information" about the EITC to refer to knowledge about the program's marginal incentive structure rather than awareness of the program's existence. Surveys of low income families and ethnographic interviews show that most EITC-eligible individuals are aware of the program's existence (as evidenced by high take-up rates), but much fewer understand the details of its structure (e.g., Ross Phillips 2001, Smeeding, Ross Phillips, and O'Connor 2002).

to report income exactly at the first kink of the EITC schedule, the point that maximizes net tax refunds (Saez 2010).[3] We show that the degree of "sharp bunching" by self-employed individuals at the first kink varies substantially across ZIP codes in the U.S. For example, 7.4% of EITC claimants in Chicago, IL are self-employed and report total earnings exactly at the refund-maximizing level, compared with 0.6% in Rapid City, SD. Bunching spreads across the U.S. and increases sharply over time: the degree of bunching is almost 3 times larger in 2009 than in 1996.

The key assumption needed to use sharp bunching as a proxy for knowledge about the EITC schedule is that individuals in low-bunching neighborhoods believe that the EITC has no impact on their marginal tax rates. We present evidence supporting this assumption in two steps. First, we show that the spatial heterogeneity in bunching is driven primarily by differences in knowledge about the first kink of the EITC schedule. We find that those who move from low-bunching to high-bunching neighborhoods are much more likely to report incomes that yield larger EITC refunds after they move. In contrast, those who move from high-bunching to low-bunching neighborhoods continue to obtain larger EITC refunds even after they move. The persistent effects of high-bunching (but not low-bunching) neighborhoods after individuals move strongly suggests that neighborhoods affect bunching via learning, as other factors would be unlikely to have such asymmetric impacts. Moreover, we find that bunching is highly correlated with predictors of information diffusion, such as the density of EITC recipients, the availability of professional tax preparers, and the frequency of Google searches for phrases including the word "tax" (e.g., "tax refund" or "Earned Income Tax Credit") in a neighborhood. In contrast, variation in local tax compliance rates or state policies explain little of the variation in bunching. Second, we show that individuals in low-bunching areas are unaware not just about the refund-maximizing kink but about the EITC schedule more broadly. In particular, when individuals become eligible for a much larger EITC refund after having their first child, the distribution of their reported self employment income remains virtually unchanged in low-bunching areas. This result establishes that individuals in low-bunching areas behave as if the EITC does not affect their marginal incentives, as required for our approach.[4]

In the second half of the paper, we use neighborhoods with low levels of sharp bunching among the self-employed (i.e., low-knowledge neighborhoods) as counterfactuals to identify the causal

---

[3]In Chetty et al. (2012), we use data from tax audits to show that this sharp bunching among the self-employed is driven primarily by non-compliance. For the analysis in this paper, it does not matter whether self-employed sharp bunching is due to manipulation of reported income or changes in real earnings.

[4]If individuals in low-bunching areas have some knowledge of the EITC schedule, our approach underestimates the impact of the EITC on earnings behavior.

impact of the EITC on the wage earnings distribution. Unlike self-employment income, wage earnings are double reported by employers to the IRS on W-2 forms. The degree of misreporting of wage earnings is therefore minimal and changes in wage earnings primarily reflect changes in "real" choices rather than non-compliance (Andreoni et al. 1998, Slemrod 2007, Chetty et al. 2012). We find that the wage earnings distribution exhibits more mass around the refund-maximizing EITC plateau in neighborhoods with high self-employed sharp bunching. Wage-earners' EITC refunds are on average 20% higher in neighborhoods in the highest sharp bunching decile relative to the lowest bunching decile. EITC refund amounts rise when wage-earners move to neighborhoods with high self-employment bunching. In contrast, moving from a high to a low bunching neighborhood does not decrease refund amounts, confirming that these effects on wage earnings are driven by learning.

The cross-neighborhood comparisons of wage earnings distributions do not definitively establish that the EITC has a causal effect on earnings because there could be other confounding differences across neighborhoods, such as differences in industrial structure or the supply of jobs. To account for omitted variable biases, we exploit the fact that individuals with no children are essentially ineligible for the EITC, thus creating a natural "control group" that can be used to account for any differences across neighborhoods that are not caused by the EITC. We implement this strategy using event studies of earnings around the birth of a first child, which effectively makes a household eligible for the EITC. The challenge in using child birth as an instrument for tax incentives is that it affects labor supply directly. We isolate the impacts of tax incentives by again using differences in knowledge about the EITC across neighborhoods to obtain counterfactuals. We find that wage earnings in low-bunching and high-bunching neighborhoods track each other closely in the years prior to child birth. However, when a first child is born, wage earnings distributions become much more concentrated around the EITC plateau in high-bunching ZIP codes, leading to larger EITC refunds in those areas. This result is robust to allowing for ZIP code level fixed effects, so that the impacts of the EITC on wage earnings are identified purely from within-area variation over time in the degree of knowledge about the schedule. Moreover, the birth of a third child – which has no impact on EITC refunds in the years we study – does not generate differential changes in earnings across areas. We conclude that unobservable differences across areas with different levels of sharp bunching are unlikely to drive our results and that the EITC has a causal impact on wage earnings.

We quantify the impacts of the EITC on average earnings behavior in the U.S. by comparing its impacts on the economy as a whole to its impacts in the lowest-bunching neighborhoods. We

find large differences between the program's impacts on earnings in the phase-in and phase-out regions. Approximately 75% of the increase in EITC refunds due to behavioral responses comes from increases in earnings in the phase-in region of the schedule, with only 25% coming from reductions in earnings in the phase-out region. The increases in EITC refunds due to behavioral responses are commensurate to a phase-in earnings elasticity of 0.14 and phase-out earnings elasticity of 0.06 on average in the U.S.. The phase-in and phase-out elasticities are 0.58 and 0.30 in the highest-knowledge areas.

One explanation for the larger responses in the phase-in is that structural labor supply elasticities are larger in the phase-in than the phase-out region, e.g. because individuals with very low incomes have higher elasticities than those holding a fixed, full time job. Another explanation is that, on average, individuals pay more attention to the phase-in and refund-maximizing plateau portions of the schedule than the phase-out region. This point illustrates a key feature of our research design: it identifies the impact of the EITC on earnings as it is *currently* perceived on average in the U.S. Changes in the structure of the program that make the phase-out incentives more salient – e.g., increasing the phase-out rate – or further diffusion of information could potentially amplify disincentive effects.

Overall, our results show that the EITC has raised net incomes at the low end of the income distribution significantly with limited work disincentive effects. The fraction of EITC-eligible wage-earners below the poverty line falls from 31.9% without the EITC to 22.0% by mechanically including EITC payments (holding earnings and reported incomes fixed). The fraction below the poverty line falls further to 21.0% once earnings responses to the EITC are taken into account. If knowledge about the EITC schedule were to increase to the level observed in the highest decile of bunching, the poverty rate would fall further to 19.6%.

Our results build on and relate to a large empirical literature on the impacts of the EITC on labor supply, surveyed by Hotz and Scholz (2003), Eissa and Hoynes (2006), and Meyer (2010). Several studies have documented clear evidence that the EITC increases labor force participation – the extensive margin response (e.g., Eissa and Liebman 1996, Meyer and Rosenbaum 2001, Eissa and Hoynes 2004, Grogger 2003, Hotz and Scholz 2006, Hotz et al. 2011, Gelber and Mitchell 2012). However, evidence on intensive margin responses is mixed and somewhat inconclusive (e.g., Meyer and Rosenbaum 1999, Bollinger et al. 2009, Rothstein 2010). The majority of the increase in EITC refunds we document here comes from individuals who were already working, providing the first non-parametric evidence that the EITC does in fact induce substantial intensive-margin responses.

Importantly, our findings do not necessarily imply that extensive margin responses are small. Our research design effectively compares areas with low vs. high levels of knowledge about the marginal incentives created by the EITC schedule. The knowledge that working can yield a large tax refund – which is all one needs to know to respond along the extensive margin – could be more widespread across all neighborhoods, perhaps because it has first-order returns (Chetty 2012).[5] But responding along the intensive margin requires knowledge about the non-linear marginal incentives created by the EITC and has only second-order benefits, potentially leading to greater variation across areas and slower diffusion over time. Our results thus help explain why prior studies of the EITC have had less success in detecting intensive margin impacts than extensive margin impacts.[6] More generally, the common wisdom that intensive margin responses to tax incentives are smaller than extensive margin responses may be an artifact of the research designs that have been used to study behavioral responses rather than a structural feature of the economy.

Our findings also contribute to the recent debate on whether EITC subsidies drive down wage rates in equilibrium, thereby limiting the extent to which the program raises net incomes (Rothstein 2010, Leigh 2010). Such general equilibrium effects are difficult to identify using traditional methods (e.g., difference-in-differences designs comparing women with and without children) because they affect both the treatment and control groups. Under the assumption that different geographic areas constitute separate labor markets, our comparisons of income distributions across neighborhoods incorporate general equilibrium changes in wage rates. Our results suggest that the EITC substantially increases earnings even when general equilibrium effects are taken into account.

Finally, our approach contributes to the recent literature on estimating the impacts of tax and transfer policies from bunching at kink points (e.g., Saez 2010, Chetty et al. 2011, Kleven and Waseem 2012) by identifying diffuse behavioral responses around kinks. Because wage-earners typically cannot control their earnings perfectly, the impact of the tax policies on the wage earnings distribution is diffuse and cannot be identified by studying the aggregate distribution. We leverage the ability to non-parametrically identify sharp bunching by self-employed tax filers through income manipulation to develop a counterfactual to identify wage-earners' diffuse real earnings responses. This method allows us to identify the impact of tax policies on the full distribution of real earnings.

---

[5] 75% of eligible individuals claim the EITC (Plueger 2009), indicating that many individuals are aware of the program's existence. This knowledge is likely due to IRS outreach efforts such as Taxpayer Assistance Centers (TAC) and Volunteer Income Tax Assistance (VITA). However, these programs focus on increasing take-up rather than disseminating information about the details of the non-linear marginal rate structure of the schedule.

[6] Liebman (1998) and Hotz and Scholz (2003, p. 182) also suggest that a lack of information could explain why the EITC has small impacts on the intensive margin.

As we discuss in the conclusion, this approach could be used to identify the impacts of a variety of policies in environments with frictions.

The remainder of the paper is organized as follows. Section II presents a stylized model to formalize our research design. Section III provides background about the EITC and the dataset we use. Section IV documents the heterogeneity across neighborhoods in sharp bunching by the self-employed and shows that this heterogeneity is driven by differences in information. Section V presents our main results on the effects of the EITC on wage earnings. In Section VI, we use our estimates to calculate the impacts of the EITC on income inequality. Section VII concludes.

II   Model and Research Design

In this section, we develop a stylized non-linear budget-set model of labor supply and tax compliance behavior to formalize our estimation strategy and identification assumptions. We make two simplifications in our baseline derivation. First, we assume that firms have constant-returns-to-scale technologies and pay workers a fixed pre-tax wage of w. Second, we abstract from income effects in labor supply by assuming that workers have quasi-linear utility functions. We discuss how these assumptions affect our estimator after analyzing the baseline case.

Setup. Individuals, indexed by i, make two choices: labor supply (l_i) and tax evasion (e_i). Let $z_i = wl_i$ denote true earnings and $\widehat{z}_i = z_i - e_i$ denote reported taxable income. Workers face a two-bracket tax system that provides a tax credit for working. When $\widehat{z}_i < K$, workers face a marginal tax rate of $\tau_1 < 0$ (a subsidy for work). For earnings above K, individuals pay a marginal tax rate of $\tau_2 > 0$ (a clawback of the subsidy). Let $\tau = (\tau_1, \tau_2)$ denote the vector of marginal tax rates.[7]

There are two types of workers: tax compliers and non-compliers. Non-compliers face zero cost of evasion and always choose to e_i to report $\widehat{z}_i = K$ and maximize their tax refunds (when they know the tax schedule, see below). Compliers face an infinite cost of altering their reported taxable income and hence always set $e_i = 0$.[8]

Individuals have quasi-linear utility functions $u(C_i, l_i, \alpha_i) = C_i - h(l_i, \alpha_i)$ over a numéraire consumption good C_i and labor supply l_i. The parameter α_i captures skill or preference heterogeneity across agents. Individuals cannot set l_i exactly at their utility-maximizing level because of frictions

---

[7]This simplifies the actual EITC schedule shown in Figure 1, which has a plateau region and two kinks. The case with one kink captures the key concepts underlying our research design.

[8]For simplicity, we ignore other variable costs of evasion, such as the threat of an audit or fines. Allowing for such costs has no impact on the estimator we derive below.

and rigidities in job packages. Our empirical approach does not rely on a specific positive model of how such frictions affect labor supply choices. Because of these frictions, the empirical distribution of true earnings $F(z)$ exhibits diffuse excess mass around the refund-maximizing kink K rather than sharp bunching at the kink K. As a result, traditional non-linear budget-set methods (e.g., Hausman 1981) and the bunching estimator proposed by Saez (2010) do not non-parametrically identify the impact of taxes on earnings behavior.

Our estimator exploits geographic heterogeneity for identification. To model such heterogeneity, we assume that there are N cities of equal size in the economy, indexed by $c = 1, ..., N$. Workers cannot move to a different city. Cities differ in their residents' knowledge about the tax credit for exogenous reasons.[9] In city c, a fraction λ_c of workers are aware of the marginal incentives τ_1 and τ_2 created by the tax credit.[10] The remainder of the workers optimize as if $\tau_1 = 0$ and $\tau_2 = 0$ (denoted below by $\tau = 0$). Cities may also differ in the distribution of skills α_i, denoted by a smooth cdf $G_c(\alpha_i)$, and in the fraction of non-compliers, θ_c. Let $F_c(z|\tau)$ denote the empirical distribution of earnings in city c with a tax system τ.

Identifying Tax Policy Impacts. Our objective is to characterize the impact of the tax credit, as it is currently perceived by agents, on the aggregate earnings distribution:

(1) $$\Delta F = F(z|\tau \neq 0) - F(z|\tau = 0).$$

The first term in this expression is the observed distribution of true earnings in the population given current knowledge of the tax credit and rates of non-compliance.[11] The second term is the potential outcome without taxes, which is the unobserved counterfactual.[12] Cities with no knowledge about the tax credit's marginal incentives ($\lambda_c = 0$) can be used to identify this counterfactual distribution. In the absence of income effects, earnings decisions in these cities are identical to behavior with no taxes at all:

$$F_c(z|\tau \neq 0, \lambda_c = 0) = F_c(z|\tau = 0, \lambda_c = 0).$$

---

[9] In practice, differences in knowledge may arise from factors related to the structure of the city, such as population density, network structure, and the availability of tax preparation services.

[10] To simplify notation, we assume that λ_c is the same for compliers and non-compliers. If knowledge varies across the types, the estimator in (3) identifies the treatment effect of interest under the two assumptions below if λ_c is interpreted as the average level of knowledge across all individuals in each city.

[11] Recall from our model that non-compliers adjust solely evasion e_i and hence their real earnings decisions z_i are not affected by knowledge about marginal tax rates. Hence, a high fraction of non-compliers would lead to attenuates real earnings responses.

[12] The traditional approach to identifying $F(z; \tau = 0|\lambda = \lambda_c)$ is to use behavior prior to a tax reform as a counterfactual. In practice, time series trends and the slow diffusion of information make it challenging to separate the causal impacts of the tax policy from confounding factors.

To use cities with $\lambda_c = 0$ as counterfactuals, we first need to measure the degree of knowledge of marginal incentives λ_c in each city. We do so by taking advantage of the fact that we observe both reported income $\widehat{z}_i$ and true wage earnings z_i in our data. The fraction of individuals in city c who report taxable income $\widehat{z}_i$ exactly at the kink, which we denote by ϕ_c, is equal to the product of local knowledge about the tax code and non-compliance rates:

$$\phi_c = \theta_c \lambda_c.$$

Hence, the rate of sharp bunching at the kink ϕ_c is a noisy proxy for the degree of knowledge λ_c. To identify areas with $\lambda_c = 0$, we make the following assumption.

Assumption 1 [Tax Knowledge]. Individuals in neighborhoods with no sharp bunching at the kink have no knowledge of the policy's marginal incentives and perceive $\tau = 0$:

$$\phi_c = 0 \Rightarrow \lambda_c = 0.$$

In our simple model, Assumption 1 is equivalent to requiring that $\theta_c > 0$ in all cities, i.e. that all cities have some non-compliers. In this case, a city with no sharp bunching at the kink must be a city in which no one knows about the tax incentives.[13] More generally, the key assumption underlying our approach is that individuals in areas with no sharp bunching behave on average as if the credit induces no change in their marginal tax rates ($\tau = 0$). If some areas with $\phi_c = 0$ actually have knowledge about marginal incentives created by the tax code, our approach will understate the impact of tax policy on earnings behavior. The degree of this attenuation bias depends on the extent to which the variation in bunching ϕ_c across cities is driven by knowledge vs. compliance rates and other factors. While we are unable to directly test Assumption 1, we present evidence that knowledge is a key driver of variation in ϕ_c and that individuals in cities with ϕ_c close to 0 behave as if they face no change in taxes ($\tau = 0$) when they become eligible for the tax credit we study.

Under Assumption 1, the empirical distribution of earnings $F_c(z)$ in cities with no sharp bunching in reported taxable income at the kink K reveals the distribution of earnings in those cities in the absence of taxes:

(2) $$F_c(z|\tau \neq 0, \phi_c = 0) = F_c(z|\tau = 0, \phi_c = 0).$$

---

[13]Importantly, Assumption 1 does *not* require that ϕ_c is an accurate proxy for differences in knowledge across all cities; it only requires when ϕ_c is low, knowledge about marginal incentives created by the tax code is low. The second requirement is much weaker and perhaps more plausible.

Although (2) identifies the necessary counterfactual in cities with no knowledge of the tax code, estimating the treatment effect in (1) requires that we identify the *mean* earnings distribution across all cities in the absence of taxes, $F(z|\tau = 0) = \frac{1}{N} \sum_{c=1}^{N} F_c(z|\tau = 0)$. This leads to the identification assumptions underlying our research design.

Assumption 2a [Cross-Sectional Identification]. Individuals' skills do not vary across cities with different levels of knowledge about the tax credit:

$$G(\alpha_i|\lambda_c) = G(\alpha_i) \text{ for all } \lambda_c.$$

This orthogonality condition guarantees that cities with low levels of sharp bunching at the kink have earnings distributions that are representative of other cities on average. Under this assumption, we obtain the following feasible non-parametric estimator for the treatment effect in (1):

$$(3) \qquad \widehat{\Delta F} = F(z|\tau) - F(z|\tau, \phi_c = 0).$$

Intuitively, the impact of the tax credit on earnings can be identified by comparing the unconditional earnings distribution with the earnings distribution in cities with no sharp bunching (i.e., no knowledge) about the tax credit.[14] Naturally, this identification strategy requires that the earnings distribution in cities with no bunching is representative of earnings distributions in other cities in the absence of taxes. We can relax this assumption by studying changes in behavior when an individual becomes eligible for the tax credit in panel data. Suppose we observe individuals making labor supply decisions for multiple years. Let t denote the year that an individual becomes eligible for the tax credit, e.g. by having a first child, which is the situation we will use in our empirical analysis. This panel design relies on a weaker "common trends" assumption for identification.

Assumption 2b [Panel Identification]. Changes in skills when an individual becomes eligible for the credit do not vary across cities with different levels of knowledge about the tax credit:

$$G_t(\alpha_i|\lambda_c) - G_{t-1}(\alpha_i|\lambda_c) = G_t(\alpha_i) - G_{t-1}(\alpha_i) \ \forall \lambda_c.$$

Under Assumption 2b, we can identify ΔF using a difference-in-differences estimator that compares earnings distributions across cities before vs. after individuals become eligible for the tax credit:

$$(4) \qquad \widehat{\Delta F}_{DD} = [F_t(z|\tau) - F_t(z|\tau, \phi_c = 0)] - [F_{t-1}(z|\tau) - F_{t-1}(z|\tau, \phi_c = 0)].$$

---

[14]In practice, there are no neighborhoods with exactly zero sharp bunching in the data. We therefore use the neighborhoods with very low levels of bunching as counterfactuals, which slightly attenuates our estimates.

The first term in (4) coincides with the cross-sectional estimator in (3). The second term nets out differences in earnings distributions across cities prior to eligibility for the credit. This estimator permits stable differences in skills across cities, but requires that skills do not trend differently across cities around the point at which individuals become eligible for the tax credit. We implement the estimator in (4) using the birth of a first child as an instrument for eligibility. Importantly, (4) permits a direct effect of child birth on labor supply as long as the effect does not differ across cities with different amounts of knowledge. Because of such direct effects, we cannot identify ΔF purely from changes in earnings behavior around the date of eligibility in the full population, again making comparisons across cities with different levels of knowledge essential for identification.[15]

Income Effects and Changes in Wage Rates. We now return to the implications of our two simplifying assumptions for our estimator for ΔF. When firms do not have constant-returns-to-scale technologies, changes in labor supply induced by tax incentives will affect equilibrium wage rates. As a result, the impact of a tax policy on the equilibrium earnings distribution is a function of both labor supply changes and changes in wage rates. The cross-sectional estimator for ΔF in (3) incorporates any such general equilibrium (GE) effects because the earnings distributions in cities with more knowledge about the tax code incorporate both changes in l_i and w_i. The difference-in-differences estimator in (4) nets out GE wage changes if individuals who are eligible and ineligible for the credit are pooled in the same market. By comparing the two estimates, one can in principle gauge the magnitude of GE effects provided that both Assumptions 2a and 2b hold.

When utility is not quasi-linear, taxes affect behavior through both price and income effects. Because individuals in all cities receive the tax credit we analyze irrespective of their perceptions, our cross-city comparisons essentially net out differences in behavior that arise purely from income effects. Hence, our estimator for ΔF approximately identifies compensated elasticities in a more general model without quasilinear utility.[16]

---

[15] In a more general model that permits heterogeneity in responses to taxation, (4) identifies the local average treatment effect of the EITC on wage earnings among households who have just had their first child.

[16] The equivalence is not exact because price effects induce changes in earnings that in turn change the size of the EITC refund that individuals in high bunching areas receive. In practice, this change in the income transfer due to behavioral responses is negligible relative to the size of the EITC and hence generates only a second-order effect.

III   Data and Institutional Background

III.A   EITC Structure

The EITC is a refundable tax credit administered through the income tax system. In 2009, the most recent year for which statistics are available, 25.9 million tax filers received a total of $57.7 billion in EITC payments (Internal Revenue Service 2011a, Table 2.5). Eligibility for the EITC depends on total earnings – wage earnings plus self-employment income – and the number of qualifying children. Qualifying dependents for EITC purposes are relatives who are under age 19 (24 for full time students) or permanently disabled, and reside with the tax filer for at least half the year.[17] Eligibility for the EITC is also limited to tax filers who are US citizens or permanent residents with a valid Social Security Number (SSN).

Figure 1a displays the EITC amount on the right y-axis as a function of earnings for single filers with one or two or more qualifying dependents throughout our period, expressed in real 2010 dollars. EITC refund amounts first increase linearly with earnings, then plateau over a short income range, and are then reduced linearly and eventually phased out completely. In the phase-in region, the subsidy rate is 34 percent for taxpayers with one child and 40 percent for taxpayers with two or more children. In the plateau (or peak) region, the EITC is constant and equal to a maximum value of $3,050 and $5,036 for filers with 1 and 2+ children, respectively. In the phase-out region, the EITC amount decreases at a rate of 15.98% for filers with 1 child, and 21.06% for those with 2+ children. The EITC is entirely phased-out at earnings equal to $35,535 and $40,363 for single filers with 1 and 2+ children, respectively. Tax filers with no dependents are eligible for a small EITC refund, with a maximum credit of $457 and a subsidy and clawback rate of 7.65%. As both the rates and levels are an order of magnitude smaller than for households with children, we exclude filers with no children from our analysis of the credit's treatment effects and use the term "EITC recipients" to refer exclusively to EITC recipients with at least one qualifying child. See IRS Publication 596 (Internal Revenue Service 2011b) for complete details on program eligibility and rules.

Aside from inflation indexation, the structure of the EITC has remained stable since 1996 after the large EITC expansion from 1994 to 1996, with two small exceptions. First, for those who are married and filing jointly, the plateau and phase-out regions of the EITC were extended by $1,000 in 2002-04, $2,000 in 2005-07, $3,000 in 2008, and $5,000 in 2009-11 (and indexed for inflation after

---

[17]Only one tax filer can claim an eligible child; for example, in the case of non-married parents, only one parent can claim the child.

2009).   Second, a slightly larger EITC was introduced for families with three or more children in 2009.   For these households, the phase-in rate is 45% (instead of 40%) with a maximum EITC of $5,666 as of 2010.   The location of the plateau remains the same as for those with two children for this group.   The stability of the EITC schedule could facilitate the diffusion of information about the program's parameters that we document below.

Note that other aspects of the tax code such as the Child Tax Credit and income taxes also affect individuals' budget sets.   Our estimates incorporate any differences across neighborhoods in knowledge about these other aspects of the tax code as well.   However, marginal tax rates in the income range we study are primarily determined by the EITC; the child tax credit and federal income tax rates have relatively small effects on incentives, as shown in Appendix Figure 1.[18]   Moreover, most of the earnings response we find comes from the phase-in region of the EITC schedule, where marginal incentives are essentially unaffected by other aspects of the tax code.   We therefore interpret our estimates as the impacts of the EITC on earnings behavior.

III.B   Sample and Variable Definitions

We use selected data from the universe of United States federal income tax returns spanning 1996-2009. Because the data start in 1996, we cannot analyze the large 1994 EITC expansion that has been used in previous work. We draw information from income tax returns (i.e., individual income tax form 1040 and its supplementary schedules) and third-party reports on wage earnings (W-2 forms). This section describes the main variables used in our empirical analysis – income, number of children, and ZIP code of residence – and the construction of our analysis samples.   In what follows, the year always refers to the tax year (i.e., the calendar year in which the income is earned). In most cases, tax returns for tax year t are filed from late January to mid-April of calendar year $t+1$. As mentioned above, we express all monetary variables in 2010 dollars, adjusting for inflation using the official IRS inflation parameters used to index the tax system.   Therefore, with the exception of the two legislated reforms described above, the EITC schedule remains unchanged in real terms across years.

Variable Definitions for Tax Filers. We use two earnings concepts in our analysis, both of which are defined at the household (tax return) level because the EITC is based on household income.   The first, *total earnings*, is the total amount of earnings used to calculate the EITC.   This

---

[18] The Child Tax Credit is only partially refundable and therefore for most of our sample period has no impact on the budget set in the phase-in region. It is quantitatively small relative to the EITC; the maximum Child Tax Credit per child is $500 before 2001 and $1,000 starting in 2001.   Federal income taxes and state income taxes typically affect the budget set starting in the phase-out region because of exemptions and deductions.

is essentially the sum of wage earnings and net self-employment earnings reported on the 1040 tax returns.[19] Total earnings correspond to reported income $\widehat{z}_i$ in our model.

The second earnings concept, *wage earnings*, is the sum of wage earnings reported on all W-2 forms filed by employers on the primary and secondary filer's behalf. Data from W-2 forms are available only from 1999 onward. For this reason, we focus primarily on the period from 1999-2009 when analyzing wage earnings impacts. However, our event studies of earnings around child birth track individuals over several years and require measures of wage earnings prior to 1999. In these cases, we define wage earnings as total wage earnings reported on the 1040 tax return form for 1996-1998.[20] We trim all income measures at -$20K and $50K to focus attention on the relevant range for the EITC.

For married individuals filing jointly, we assign both individuals in the couple the household-level total earnings and wage earnings because the EITC is based on household income. However, we structure our analysis based on an individual-level panel to account for potential changes in marital status. Because we define earnings at the family level, changes in marital status can affect an individual's earnings even if his or her own earnings do not change.[21]

We define the number of children as the number of children claimed for EITC purposes. The EITC children variable is capped at 2 from 1996-2007 and 3 in 2009. For individuals who report the maximum number of EITC children, we define the number of children as the maximum of EITC children and the number of dependent children claimed on the tax return. If the number of children claimed for EITC purposes is missing because the tax return does not claim the EITC (e.g., because earnings are above the eligibility cutoff), we define the number of children as the number of dependent children.[22]

Finally, we define ZIP code as the ZIP code from which the individual filed his year t tax return. If an individual did not file in a given tax year, then we use the ZIP code reported as the home

---

[19] More precisely, total earnings is the sum of the wage earnings line on the 1040 plus the Schedule C net income line on the 1040 form minus 1/2 of the self-employment tax on the 1040 adjustments to gross income. This adjustment is made in the tax code to align the tax treatment of wage earnings and self-employment earnings for Social Security and Medicare taxes. These taxes are split between employers and employees for wage earners, and wage earnings are reported net of the employer portion of the tax.

[20] Total wage earnings reported on the tax return also include some minor forms of wage earnings not reported on W-2 forms, such as tips. The W-2 earnings measure is preferable because individuals could misreport wage income that is not third party reported on W-2 forms. None of our results are sensitive to the exclusion of pre-1999 data because we only use these data to assess pre-period trends, as discussed in greater detail below.

[21] We have checked that our results are not driven by marriage effects by re-doing the analysis using solely individual earnings, instead of family earnings.

[22] The requirements for EITC-eligible children vs. dependent children are not identical, but the difference is minor in practice. According to our calculations from the 2005 Statistics of Income Public Use Microdata File, less than 10% of EITC filers report different numbers for dependent children and EITC children.

address on the W-2 with the largest earnings reported for that individual in that year.

We do not observe total earnings or number of children for individuals who do not file tax returns, and we do not observe ZIP code for individuals who neither file nor earn wages reported on a W-2. These missing data problems can potentially create selection bias, which we address in our child birth sample below.

Core Sample. Our analysis sample includes individuals who meet all three of the following conditions simultaneously in at least one year between 1996 and 2009: (1) file a tax return as a primary or secondary filer (in the case of married joint filers), (2) have total earnings below $50,000 (in 2010 dollars), and (3) claim at least one child. We impose these restrictions to limit the sample to individuals who are likely to be EITC-eligible at least once between 1996 and 2009. We also remove observations with ITINs from the sample.[23] We define the total earnings and wage earnings of person-year observations with no reported earnings activity as zero. These include individuals who do not file a tax return and have no W-2 wage earnings, individuals who die within the sample period, and individuals who leave the United States. This procedure yields a balanced panel with no attrition, i.e. every individual has exactly fourteen years of data. We refer to the resulting sample as our core analysis sample. The core sample contains 77.6 million unique individuals and 1.09 billion person-year observations on earnings. Our empirical analysis consists of three different research designs, each of which uses a different subsample of this core sample.

Cross-Sectional Analysis Sample. Our first research design compares earnings distributions for EITC claimants across cities in repeated cross-sections. For this cross-sectional analysis, we limit the core sample to person-years in which the individual files a tax return, reports one or more children, has total earnings in the EITC-eligible range, and is the primary filer. By including only primary filers, we eliminate duplicate observations for married joint filers and obtain distributions of earnings that are weighted at the tax return (family) level, which is the relevant weighting for tax policy and revenue analysis. Note that this cross-sectional sample excludes non-filers and thus could in principle yield biased results if EITC take-up rates vary endogenously across cities. We cannot resolve this problem in cross-sections because we do not observe non-filers' number of children. We therefore address this issue using panel data in our third research design below.

Movers Sample. Our second research design tracks individuals as they move across neighborhoods. To construct the sample for this analysis, we first limit the core sample to person-years in

---

[23]The IRS issues ITINs to individuals who are not eligible for a Social Security Number (and are thus ineligible for the EITC). These individuals include undocumented aliens and temporary US residents, and account for 2.6% of our core sample.

which an individual files a tax return, claims one or more children, and has income in the EITC-eligible range.[24] We then further restrict the data to individuals who move across 3-digit ZIP codes (ZIP-3s) in some year between 2000 and 2005.[25] We impose these restrictions to ensure that we have at least four years of data on earnings before and after the move. In addition, this restriction also guarantees that we have W-2 (employer reported) wage earnings data for at least one year before the move. We define a move as a change in ZIP-3 between two consecutive years for which address information is available. When individuals move more than once, we include only the first move (as well as 4 years on either side, regardless of the timing of the second move). Note that we observe address at the time of tax filing, which in the EITC population is typically February of year $t + 1$ for year t incomes. A change in address for tax year t therefore implies that the move most likely took place between February of year t and February of year $t + 1$. A small fraction of the moves classified as occurring in year t thus do not take place till shortly after the end of that year. Importantly, none of the moves classified as occurring in year t occur prior to year t with this definition, ensuring that any misclassification errors do not affect pre-move distribution and only attenuate post-move impacts.

Child Birth Sample. Our third research design tracks individuals around the year in which they have a child, which can trigger eligibility for a larger EITC.[26] We observe dates of birth as recorded by the Social Security administration. As in the movers sample, we restrict attention to births between 2000 and 2005 to ensure that we have at least 4 years of earnings data before and after child birth and at least one year of pre-birth W-2 earnings data. Next, we define the parents of the child as all the primary and secondary filers that claim the child either as a dependent or for EITC purposes within 5 years of the child's birth. If the child is claimed by multiple individuals (e.g., a mother and father filing jointly), we define both individuals as new parents and track both parents over time. We then limit the core sample to the set of all such new parents, including all observations regardless of whether the individuals files a tax return in a given year.

In our child birth sample, we impute non-filers' earnings, addresses, and number of children as follows. Because marital status is only observed on income tax forms, we cannot identify spouses for non-filers. We assume that non-filers are single and define both their total earnings and wage

---

[24] We include both primary and secondary filers to avoid excluding a subset of observations for individuals who change marital status within our sample. We account for repeated observations for married joint filers by clustering standard errors as described below.

[25] See Section IV.B below for a detailed description of ZIP-3s.

[26] As in the movers sample, we include all individuals (both primary and secondary filers) rather than families here to avoid dropping observations when marital status changes.

earnings as the total income reported on W-2 forms.[27] We code total earnings as zero for non-filers who have no W-2's.[28] Throughout the sample, we assign individuals the ZIP code in which they lived during the year in which the child was born. For non-filers, we impute the ZIP code as the ZIP code to which a W-2 form was mailed in the year of child birth if available.[29] 11.6% of households neither filed a tax return nor had W-2 information in the year their child was born; for this group we use the first available ZIP code after the child was born. Finally, we impute the number of children for non-filers as the minimum of the children claimed in the closest preceding and subsequent years in which the individual filed (not including the child who was born in year 0).[30]

With these imputations for non-filers, the child birth sample includes all years for every individual who (1) has a child born between 2000 and 2005 according to Social Security records and (2) claims that child on a tax return at some point after his birth. Treating the decision to have a child as exogenous, the only selection into this child birth sample comes from the potentially endogenous decision to claim a child as a dependent.[31] We account for potential selection bias through this channel using data prior to child birth as described below.

Descriptive Statistics. Table I presents summary statistics for our cross-sectional analysis sample using data from 1999-2009, the years in which we have W-2 earnings information. Mean total earnings are $20,091. The majority of this income comes from wage earnings: mean wage earnings as reported on W-2's are $18,308. 19.6% of tax filers report non-zero self-employment income and the mean (unconditional) self-employment income in the sample is $1,770. Individuals in this population receive substantial EITC refunds, with a mean of $2,543. Nearly 70% of the tax returns are filed by a professional preparer. The population of EITC eligible individuals consists primarily of relatively young single women with children.

---

[27]Excluding elderly households who receive Social Security Income, over 90% of non-filers are single (Cilke 1998, Table 1, p. 15). Because our sample requires having a child birth at some point within the sample, it contains very few elderly households. Self-employment earnings are not observed if the individual does not file and are assumed to be zero.

[28]This procedure codes total earnings and wage earnings as 0 for non-filers prior to 1999, when W-2 data are unavailable. Most non-filers have very low W-2 earnings when data are available, so this imputation is likely to be accurate for most cases. As noted above, none of our results are sensitive to the exclusion of pre-1999 data.

[29]For individuals with multiple W-2 forms, we use the W-2 with the largest amount of earnings and non-missing address information.

[30]While we cannot be certain about the number of dependents living with an individual in years she does not file, it is more likely that the number of children is the minimum of the lead and lag as children are sometimes exchanged (for tax reporting purposes) across parents. Individuals who do not file are therefore likely to have fewer children.

[31]The empirical literature on the EITC has found no evidence that the EITC affects marriage and fertility decisions (Hotz and Scholz 2003, p. 184).

IV   Neighborhood Variation in Bunching and EITC Knowledge

In this section, we develop a proxy for local knowledge about the EITC in four steps. First, we document sharp bunching at the first kink of the EITC schedule by self-employed individuals in the aggregate income distribution. Second, we show that the degree of sharp bunching varies substantially across neighborhoods in the U.S. Third, we present evidence that this spatial variation is driven by differences in knowledge about the refund-maximizing kink of the EITC schedule rather than other factors such as local tax compliance rates. Finally, we show that individuals in low-bunching areas are unaware not only of the refund-maximizing kink but behave as if the EITC does not affect their marginal tax rates at all income levels. Together, the results in this section establish that self-employed sharp bunching is a proxy for local knowledge that satisfies Assumption 1 above.

IV.A   Aggregate Distributions: Self-Employed vs. Wage Earners

Figure 1a plots the distribution of total earnings for EITC claimants in 2008 using our cross-sectional analysis sample. The distribution is a histogram with $1,000 bins centered around the first kink of the EITC schedule. We plot separate distributions for EITC filers with one and two or more children, as these individuals face different EITC schedules, shown by the solid lines in the figures.[32] Both distributions exhibit sharp bunching at the first kink point of the corresponding EITC schedule, the point that maximizes tax refunds net of other income tax liabilities (such as payroll taxes). This sharp bunching shows that the EITC induces significant changes in reported income, confirming Saez's (2010) findings using public use samples.

Figure 1b replicates Figure 1a restricting the sample to wage-earners, defined as households who report zero self-employment income in a given year. In this figure, there is no sharp bunching at the EITC kinks, implying that all the sharp bunching in Figure 1a is due to the self-employed. However, one cannot determine from Figure 1b whether the EITC has an impact on the wage earnings distribution. The impact for wage-earners is likely to be much more diffuse because they cannot control their earnings perfectly due to frictions (Chetty et al. 2011). One therefore needs counterfactuals for the distributions in Figure 1b to identify the impacts of the EITC on wage earnings. We show below that the wage earnings distributions in Figure 1b are in fact reshaped by the EITC, but one would have no way of detecting such responses by studying only the aggregate

---

[32]These and subsequent figures include both single and married individuals. Married individuals face an EITC schedule with a slightly longer plateau region but the same first kink point. The EITC schedules shown in Figure 1 are for single individuals.

distribution.[33]

We develop counterfactuals for the wage earnings distribution using the research design described in Section II. To implement the approach empirically, we interpret the sharp bunching among the self-employed as a measure of manipulation in total earnings ($\hat{z}_i$ in the model) and wage earnings reported on W-2's as true earnings (z_i). Because wage earnings are double reported by employers to the IRS through W-2 forms, individuals have little scope to misreport wage earnings.[34] In contrast, there is no systematic third-party reporting system for self-employment income and the expenses and profits of small businesses are difficult to verify, making it much easier to misreport self-employment income. Random audits reveal substantial misreporting of income among self-employed individuals, whereas compliance rates for wage earnings exceed 98% (Internal Revenue Service 1996).[35] In a companion paper (Chetty et al. 2012), we replicate this finding within the EITC population using audit data from the 2001 National Research Program. We find that the majority of the sharp bunching at the first kink of the EITC schedule among the self-employed is due to non-compliance, as the degree of sharp bunching in the post-audit total earnings distribution falls to 1/3 of the original level. In contrast, misreporting among wage-earners is negligible even around the refund-maximizing region of the schedule, supporting the view that wage earnings represent true earnings z_i.

In the remainder of this section, we focus on total earnings ($\hat{z}_i$) and analyze variation across neighborhoods in the degree of self-employed sharp bunching at the first kink of the EITC schedule.

IV.B Spatial Heterogeneity in Sharp Bunching

We analyze spatial heterogeneity at the level of three-digit ZIP codes, which we refer to as ZIP-3s.[36] We define the degree of self-employed sharp bunching in a ZIP-3 as the percentage of EITC claimants who report total earnings at the first EITC kink and have non-zero self employment

---

[33] There is no need for a counterfactual to estimate sharp bunching among the self-employed because there is no reason to expect point masses in the income distribution at the kinks of the tax schedule except for the impact of the tax system itself. By leveraging our ability to non-parametrically identify sharp bunching without a counterfactual, we develop counterfactuals to identify the diffuse response of wage earners.

[34] Any discrepancy between an individual tax return self-report and the employer W2 information return report is automatically detected by the IRS and can trigger an audit. Misreporting wage earnings therefore requires collusion between employers and employees, which is likely to be difficult especially in large firms. We show below that our results hold in the subsample of wage earners working at firms with more than 100 employees.

[35] For instance, the rate of income under-reporting for small business suppliers was over 80 percent in 1992 (Internal Revenue Service 1996, Table 3, page 8).

[36] Standard (5 digit) ZIP codes are typically too small to obtain precise estimates of income distributions. Common measures of broader geographical areas such as counties or MSA's are more cumbersome to construct in the tax data or do not cover all areas. There are 899 ZIP-3s in use in the continental United States, shown by the boundaries in Figure 2. ZIP-3 are typically (but not always) contiguous and are smaller in dense areas. For example, in Boston, the 021 ZIP-3 covers roughly the same area as the metro area's subway system.

income. More precisely, for ZIP-3 c in year t, let num_{ct} denote the number of primary tax filers who claim the EITC with children, report non-zero self employment income, and report total earnings within \$500 of the first kink. Let denom_{ct} denote the total number of primary tax filers with children in ZIP-3 c in year t in our cross-sectional analysis sample. We define self-employed sharp bunching b_{ct} as $\text{num}_{ct}/\text{denom}_{ct}$. Note that this definition incorporates both intensive and extensive margin changes in reporting self-employment income. Thus, part of the variation in bunching across areas is driven by differences in rates of reporting self-employment income, some of which is endogenous to knowledge about the EITC as we show below.[37]

Figure 2 illustrates the spatial variation in b_{ct} in 2008 across the 899 ZIP-3s in the United States. To construct this figure, we divide the raw individual-level cross sectional data in 2008 into 10 deciles based on b_{ct}, so that the deciles are population-weighted rather than ZIP-3 weighted.[38] Higher deciles are represented with darker shades on the map. The mean (population weighted) level of b_{ct} in the U.S. in 2008 is 2.4%. To gauge magnitudes, recall that the mean self-employment rate in our sample is approximately 20%; hence, if 10% of self-employed EITC claimants report total earnings at the kink, we would observe $b_{ct} = 2\%$.

There is substantial dispersion in self-employed sharp bunching across neighborhoods in the U.S.[39] For example, bunching rates are less than 0.5% in most parts of North and South Dakota, but are over 5% in some parts of Texas and Florida. While some of the variation in bunching occurs at a broad regional level – for example, bunching is greater in the Southern states – there is considerable variation even within nearby geographical areas. For example, the Rio Grande Valley in Southern Texas has self-employed sharp bunching of $b_{ct} = 6.6\%$; in contrast, Corpus Christi, TX, which is 150 miles away, has bunching of $b_{ct} = 2.3\%$. Moreover, there are no obvious discontinuities at state borders, suggesting that differences in state policies such as state EITC's are unlikely to explain the heterogeneity, a result that we verify formally below.

Appendix Figure 3 replicates Figure 2 for earlier years, beginning in 1996, the first year of our dataset and the year in which the EITC was expanded to its current form. To illustrate variation

---

[37] We have assessed the robustness of our results to several alternative measures of sharp bunching, including (a) defining the denominator using only self-employed individuals rather than the full population to eliminate variation arising from differences in self-employment rates; (b) defining narrower and wider bands than \$500 around the kink; and (c) calculating excess mass relative to a smooth polynomial fit as in Chetty et al. (2011). Because self-employed bunching is so sharp (as shown in Figure 1), our results are essentially unchanged with these alternative definitions. As an illustration, we replicate our main results using the definition in (a) in Appendix Figure 2.

[38] Visually, most of the country appears to be in the lower bunching deciles because bunching rates are much higher in dense neighborhoods, as we show below.

[39] Given the sample sizes – which are on average 23,000 returns per ZIP-3 – bunching rates are essentially estimated without error and we therefore ignore the impact of imprecision in our estimates of b_{ct}.

over time, we divide the observations into deciles after pooling all years of the sample, so that the decile cut points remain fixed across years. Initially, sharp bunching was highly prevalent in a few areas with a high density of EITC filers, such as southern Texas, New York City, and Miami. Bunching has since spread throughout much of the United States and continues to rise over time.

Figure 3 plots the distribution of total earnings for individuals living in the lowest and highest bunching deciles in the pooled sample from 1996-2009. This figure includes individuals with both 1 and 2+ children by plotting total earnings minus the first kink point of the relevant EITC schedule, so that 0 denotes the refund-maximizing point. In the top decile, more than 8% of tax filers report total earnings exactly at the refund-maximizing kink. In contrast, there is virtually no bunching at this point in neighborhoods in the bottom decile, suggesting that these neighborhoods could provide a good counterfactual for behavior in the absence of the EITC if the lack of sharp bunching is due to a lack of knowledge about the EITC schedule.

IV.C   Is the Variation in Bunching Driven by Knowledge?

We evaluate whether the differences in self-employed sharp bunching across ZIP-3s are driven by differences in knowledge about the refund-maximizing kink of the schedule using two tests. First, we analyze individuals who move across ZIP-3s and test for learning. Second, we correlate bunching rates with proxies for the rate of information diffusion and competing explanatory factors such as tax compliance rates.

Movers. Our hypothesis that the variation in bunching is driven by differences in knowledge generates two testable predictions about the behavior of movers. The first is learning: individuals who move to a higher bunching area should learn from their neighbors and begin to respond to the EITC themselves. The second is memory: individuals who leave high bunching areas should continue to respond to the EITC even after they move to a lower bunching area. This asymmetric impact of prior and current neighborhoods distinguishes knowledge from other explanations for heterogeneity in bunching. For instance, variation in preferences or tax compliance rates across areas do not directly predict that an individual's previous neighborhood should have an asymmetric impact on current behavior.

We implement these two tests using the movers sample defined in Section III, which includes all individuals in our core sample who move across ZIP-3s at some point between 2000 and 2005. This sample includes 21.9 million unique individuals and 54 million observations spanning 1996-2009. We define the degree of bunching for prior residents of ZIP-3 c in year t as the sharp bunching

20

rate for individuals in the cross-sectional analysis sample living in ZIP-3 c in year $t - 1$. We then divide the ZIP-3-by-year cells into deciles of prior residents' bunching rates by splitting the individual-level observations in the movers sample into ten equal-sized groups. Note that with this definition, ZIP-3s may change deciles over time if their bunching rates rise or fall.

Figure 4a plots an event study of bunching for movers around the year in which they move. To construct this figure, we first define the year of the move as the first year a tax return was filed from the new ZIP-3. We then compute event time as the calendar year minus the year of the move, so that event year 0 is the first year the individual lives in the new ZIP-3. For illustrative purposes, we focus on individuals who live in a ZIP-3 in the fifth decile of the overall bunching distribution in the year prior to the move. We then divide this sample into three groups based on where they move in year 0: the first, fifth, and tenth bunching deciles. We calculate the sharp bunching rate in each event year and subgroup as the fraction of EITC claimants in the relevant group who report total earnings at the first kink and have non-zero self employment income.

To obtain a point estimate of the effect of moving to decile 10, we regress an indicator for sharp bunching (i.e., reporting total earnings at the kink and non-zero self employment income) on an indicator for moving to decile 10, an indicator for event year 0, and the interaction of the two indicators. We estimate this regression restricting the sample to event years -1 and 0 and deciles 5 and 10, so that the coefficient on the interaction term (β_{10}) is a difference-in-differences estimate of the impact of moving to decile 10 relative to decile 5. We estimate treatment effects of moving to deciles 1 and 5 using analogous specifications, always using decile 5 as the control group. Standard errors, reported in Figure 4 in parentheses below the coefficient, are clustered at the destination ZIP-3-by-year-of-move level.

Bunching rates rise sharply by $\beta_{10} = 1.9$ percentage points for individuals who move to the highest bunching decile, rise by a statistically insignificant $\beta_5 = 0.1$ percentage points for those who stay in a fifth-decile area, and fall slightly (by $\beta_1 = -0.4$ percentage points) for those who move to the lowest bunching decile. Individuals rapidly adopt local behavior when moving to high bunching areas. The mean difference in self-employed sharp bunching rates for prior residents is 3.6 percentage points between the fifth and tenth deciles. Hence, movers to the top decile adopt $(2.0-.1)/3.6 = 53\%$ of the difference in prior residents' behavior within the first year of their move.

While sharp bunching is perhaps the clearest evidence of responding to the EITC, relatively few individuals report income exactly at the first kink. To evaluate whether individuals learn about the EITC schedule more broadly when they move, we plot mean EITC refunds by event

year in Figure 4b. Using a difference-in-differences specification analogous to that used in Figure 4a, we estimate that EITC refund amounts rise by $150 on average when individuals move to the highest bunching decile. The increase in sharp bunching at the first kink accounts for at most $1.9\% \times \$4,403 = \$77$ of this increase.[40] Hence, individuals report incomes that generate larger EITC refunds more broadly than just around the first kink when they move to areas with high levels of sharp bunching.

Figure 5 plots total earnings distributions in the years before and after the move for the three groups in Figure 4. This figure is constructed in the same way as Figure 3, pooling individuals with 1 and 2+ children and computing total earnings relative to the first kink of the relevant EITC schedule. Consistent with the results from the event studies, the fraction of individuals reporting total earnings exactly at the kink and around the refund-maximizing plateau increases significantly after the move for those moving to high bunching areas, consistent with learning.[41] However, the distribution remains relatively stable for those moving to low bunching areas, consistent with memory.

To distinguish learning and memory more directly, we test for asymmetry in the impacts of increases vs. decreases in sharp bunching rates when individuals move. Figure 6 plots changes in mean EITC refunds from the year before the move (year -1) to the year after the move (year 0) vs. the change in local sharp bunching Δb_{ct} that an individual experiences when he moves. Following standard practice in non-parametric regression kink designs, we bin the x-axis variable Δb_{ct} into intervals of width 0.05% and plot the means of the change in EITC refund within each bin. If the variation in bunching is due to knowledge, there should be a kink in this relationship around 0: increases in b_{ct} should raise refunds, but reductions in b_{ct} should leave refunds unaffected. We test for the presence of such a kink by fitting separate linear control functions to the points on the left and right of the vertical line, with standard errors clustered by the bins of Δb_{ct} (Card and Lee 2007). As predicted, the slope to the right of the kink is significant and positive: a 1 percentage point increase in sharp bunching at $b_{ct} = 0$ leads to a $60 increase in EITC refunds. In contrast, a 1 percentage point reduction in b_{ct} leads to a statistically insignificant change in EITC refunds

---

[40]Roughly half of the individuals in the movers sample claim one child, while the other half claim two or more children. The weighted average of the maximum EITC refund across these groups is $4,043. $77 is a non-parametric upper bound on the impact of sharp-bunching on average EITC refunds; the actual effect is likely much smaller.

[41]Individuals moving to decile 10 exhibit more bunching even prior to the move because our ZIP-3 measure of neighborhoods generates discrete jumps in neighborhood bunching at boundaries. Individuals who move to decile 10 are more likely to live in ZIP-3's that are adjacent to decile 10 areas, and thus live in locally higher bunching areas even though their ZIP-3 is classified in decile 5 as a whole. This measurement error in neighborhood bunching works against the hypotheses we test.

of $6. The hypothesis that the two slopes are equal is rejected with $p < 0.0001$. The kink at zero constitutes non-parametric evidence of asymmetric responses to changes in bunching rates and therefore strongly indicates that at least part of the variation in b_{ct} is due to knowledge.[42]

Cross-Sectional Correlations. To better understand the sources of variation in sharp bunching, we correlate b_{ct} with proxies for information, tax compliance, and other variables. While we cannot interpret these correlations as causal effects, the relative explanatory power of various factors sheds some light on why knowledge varies so much across areas.

Table II presents a set of OLS regressions of the rate of sharp bunching in each ZIP-3 on various correlates. Among a broad range of economic and demographic variables available from the 2000 decennial Census, the single strongest predictor of sharp bunching is the local density of EITC filers. In column 1 of Table II, we regress sharp bunching on density of EITC filers, defined as the number of EITC claimants with children (measured in 1000's) per square mile. We estimate the regression in a dataset that has one observation on sharp bunching per ZIP-3 in 2000 (the year of the Census) and weight by the number of EITC claimants in each ZIP-3. Increasing the density of EITC filers by 1,000 per square mile (a 1.6 SD increase) raises bunching rates by 1.93 percentage points (1.1 SD). The R-squared of the density variable by itself in a univariate regression (weighted by the number of filers in each ZIP-3) is 0.6. Intuitively, this regression shows that an isolated EITC recipient is less likely to learn about the schedule than one living amongst many other EITC eligible families.

The correlation between density and sharp bunching suggests that agglomeration facilitates the diffusion of knowledge in dense areas. Figure 7a documents this diffusion over time by plotting the average level of sharp bunching by year from 1996-2009. We split the sample into two groups: ZIP-3s with EITC filer density below vs. above the median in 1996. The degree of sharp bunching was relatively similar across these areas in 1996, the first year of the current EITC schedule. But rates of bunching rose much more rapidly in dense areas, presumably because information about the EITC schedule diffused more quickly in these areas.

Column 2 of Table II adds the following additional demographic controls to the specification in column 1: the percentage of the population that is foreign born, white, black, Hispanic, Asian, and other race. Bertrand et al. (2001) suggest that these demographic characteristics are related to the tightness of networks in low income populations. Consistent with this hypothesis, we find that

---

[42]We show below that wage earnings exhibits similar asymmetric persistence, implying that individuals learn not just about non-compliance but also about the incentives that affect real work decisions.

these demographic characteristics explain a substantial share of the variation in sharp bunching beyond density, increasing the R-squared from 0.6 to 0.8.

Prior studies have also suggested that professional tax preparers may help disseminate information about the tax code (e.g., Maag 2005, Chetty and Saez 2012). To evaluate this hypothesis, in column 3 of Table II, we regress sharp bunching on the fraction of individuals who use a tax preparer in each ZIP-3 of our cross-sectional analysis sample in 2008. A 10 percentage point (1.5 SD) increase in the rate of local tax professional utilization is associated with a 0.986 percentage point (0.57 SD) increase in sharp bunching. Figure 7b plots the relationship between sharp bunching and the fraction of professionally prepared returns in the ZIP-3, dividing claimants into two groups based on whether they themselves used a tax preparer or not. This figure is a binned scatter plot, constructed by binning the x-axis into 20 equal-sized bins (vingtiles) and plotting the means of b_{ct} for each group in each bin. The figure shows that areas with high tax preparer penetration exhibit higher bunching among *both* groups. This result implies that tax professionals either serve simply as a seed for knowledge – informing their clients about the EITC who in turn spread the information to others – or that tax preparation firms locate endogenously in areas where EITC refunds are already high (Kopczuk and Pop-Eleches 2007).

Column 4 of Table II shows that sharp bunching is highly correlated with Google searches for information about taxes and tax refunds, another proxy for interest in and awareness about tax-related information.[43] Following the techniques developed by Stephens-Davidowitz (2011), we measure the percentage of an area's Google searches for any phrase that includes the word "tax" (such as "Earned Income Tax Credit" or "tax refund") between 2004 and 2008. We divide this measure by its standard deviation to obtain a standardized measure.[44] We regress sharp bunching on the Google search measure using the cross-sectional analysis sample in 2008, as internet usage rates were much lower in 2000 than 2008. A 1 SD increase in Google search intensity for "tax" in a ZIP-3 is associated with an 0.3 percentage point (0.17 SD) increase in sharp bunching. This association remains statistically significant when we add demographics, density, and professional tax preparation rates to the specification, as shown in column 5. Column 6 replicates column 5 with state fixed effects. EITC filer density, tax preparation services, and searches for information

---

[43]Internet usage is substantial even amongst low SES populations: according to data from the CPS, 39% of individuals who did not graduate high school lived in a household with internet access in 2009 (U.S. Census Bureau 2012). We use the search term "tax" rather than more specific terms such as "EITC" because many individuals may not know the term "EITC" and because the Google search statistics are publicly available only for words that appear in a large number of searches.

[44]Google search data are obtained at a media market level, which we map to ZIP-3's using population-weighted averages.

about taxes remain highly predictive of within-state variation in sharp bunching.

Finally, we evaluate some competing explanations for the spatial variation in bunching. Column 7 shows that differences in state EITC top-up rates do not have a statistically significant impact on sharp bunching rates and explain relatively little of the variation in bunching. In column 8, we analyze whether differences in tax compliance rates (θ_c) across areas explain the variation in sharp bunching. We implement this analysis using data on random audits from the 2001 National Research Program as follows.[45] First, we define a measure of non-compliance in each state as the fraction of non-EITC claimants who have adjustments of more than \$1000 in their income due to NRP audits. We define non-compliance rates using individuals who do not receive the EITC to eliminate the mechanical correlation arising from the fact that individuals bunch at the kink primarily by misreporting total earnings. We then regress sharp bunching among EITC claimants in each state on the non-compliance rate, weighting by the number of individuals audited in each state to adjust for differences in sampling weights in the NRP. The correlation between sharp bunching and non-compliance rates is statistically insignificant, as shown in column 8. The non-compliance measure has an R-squared of less than 1% by itself, suggesting that spatial variation in bunching is unlikely to be driven by heterogeneity in non-compliance.

In sum, the correlations indicate that a substantial fraction of the variation in sharp bunching across areas reflects differences in knowledge about the refund-maximizing kink of the EITC schedule that arise from structural features of local economies such as population density and demographic characteristics.

IV.D Perceptions of the EITC in Low-Bunching Areas

While the preceding evidence establishes that self-employed sharp bunching provides an informative (albeit noisy) proxy for local knowledge about the first kink of the EITC schedule, it does not directly establish that Assumption 1 holds. For instance, individuals who live in low-bunching areas may perceive the EITC to be a flat subsidy at a constant rate or a smoothly varying subsidy without kinks in the schedule. Such misperceptions would generate no bunching at the first kink but would imply that low-bunching areas do not provide a valid counterfactual for behavior in the absence of the EITC. We now present evidence that individuals in low-bunching areas actually appear to have no knowledge about the entire EITC schedule and behave as if $\tau = 0$ on average

---

[45] State-level tabulations from NRP data were provided by the IRS Office of Research. Note that the NRP sampling frame was not explicitly designed to be representative at the state level, so the results here should be interpreted with caution.

when they become eligible for the credit.

We assess the beliefs of individuals in the lowest-bunching decile by examining changes in the distribution of reported self-employment income around the birth of a first child. As noted above, this event makes families eligible for a much larger EITC refund and sharply changes marginal incentives. We implement this analysis using our child birth sample, which includes approximately 15 million individuals from the core sample who have their first child between 2000 and 2005. We classify individuals into deciles of sharp bunching based on the level of b_{ct}, as measured from the cross-sectional sample, in the ZIP-3 and year in which he or she had a child.

Figure 8a plots the distribution of total earnings among self-employed individuals in the year before birth and the year of child birth. The distributions are scaled to integrate to the total fraction of individuals reporting self-employment income in each group, which varies across the groups as shown in Figure 8b below. The reported earnings distribution changes only slightly when individuals in the lowest-bunching decile have a child. In contrast, the distribution of total reported income exhibits substantial concentration both at and around the first kink for individuals in the top-bunching decile.[46] The fact that the total earnings distribution remains virtually unchanged when individuals have a child in low-bunching areas implies that they perceive no changes in marginal incentives throughout the range of the EITC (rather than simply ignoring the first kink). For instance, if individuals in low-bunching areas perceived the EITC to be a constant subsidy, we would observe an upward shift in the total reported income distribution when individuals have a child and become eligible for the EITC.

Figure 8b conducts an analogous test on the extensive margin by plotting the fraction of individuals reporting self employment income by event year around child birth, which is denoted by year 0. While there are clear trend breaks in the fraction reporting self-employment income around child birth in higher-bunching areas, there is little or no break around child birth in the lowest-bunching decile. Although we have no counterfactual for how self-employment income would have changed around child birth in low-bunching areas absent the EITC, we believe that the costs of manipulating reported self-employment income are unlikely to change sharply around child birth.[47] Hence, the smooth trends in self-employment rates around child birth in the lowest-decile bunching areas

---

[46]To simplify the figure, we only plot the distribution of earnings in the year before the birth for households in low-bunching neighborhoods. The pre-birth distribution in high bunching areas is similar to that in low-bunching areas; in particular, it does not exibit any sharp bunching around the first kink of the EITC schedule.

[47]Recall that the audit evidence reveals that changes in self-employment income are largely driven by non-compliance and hence reflect pure reporting effects. In contrast, child birth clearly has effects on real labor supply, making it crucial to have a counterfactual when using child birth as a quasi-experiment to identify wage earnings impacts as we do in Section V below.

provide further evidence that individuals in these areas do not perceive any change in incentives when they have a child.

Provided that individuals perceive $\tau = 0$ before they are eligible for the EITC, the results in Figure 8 imply that EITC-eligible individuals in low bunching areas perceive and behave as if $\tau = 0$ on average, as required by Assumption 1.[48] We therefore proceed to use low-bunching neighborhoods as counterfactuals for behavior in the absence of the EITC. Note that we would ideally use areas with literally zero bunching as counterfactuals. In practice, there are very few areas with literally no sharp bunching, but the level of sharp bunching is very close to zero in the bottom decile, as shown in Figures 3 and 8a. We therefore use the lowest bunching decile as "no knowledge" areas to avoid extrapolations and maintain adequate precision to estimate counterfactual distributions. Our estimates slightly understate the causal impacts of the EITC because of this simplification.

V   Effects of the EITC on Wage Earnings

In this section, we identify the impacts of the EITC on the distribution of real wage earnings using self-employed sharp bunching as a proxy for local knowledge about the EITC. We present estimates from two research designs. We first compare earnings distributions across neighborhoods in cross-sections. We then use child birth as a source of sharp changes in marginal incentives to obtain estimates from panel data that rely on weaker identification assumptions.

Throughout most of this section, we limit the sample to wage-earners (individuals who report zero self-employment income) and analyze wage earnings as reported by firms on W-2 forms. Note that restricting the sample based on self-employment income could in principle introduce selection bias, as the choice to report self-employment income is endogenous and depends upon knowledge about the EITC. In the last part of this section, we show that including all individuals and using W-2 wage earnings as the outcome yields similar but less precisely estimated results, implying that endogenous selection is not a significant source of bias in practice.

V.A   Cross-Neighborhood Comparisons

We begin by comparing the distribution of wage earnings in ZIP-3s with low vs. high levels of sharp bunching. Identifying the causal impacts of the EITC using this research design requires

---

[48]Individuals in the EITC income range who do not have children pay minimal taxes and receive minimal refunds; hence, it is most plausible that they perceive essentially zero marginal tax rates. These individuals may be aware of some aspects of the tax schedule, such as payroll or income taxes. In that case, our approach would identify the impact of the tax system including the EITC as it is perceived in the population on average relative to tax perceptions absent the EITC.

that areas with different levels of sharp bunching would have comparable earnings distributions in the absence of the EITC (Assumption 2a). In practice, there could be many differences across ZIP-3s with different levels of sharp bunching, as they differ in population density and various other characteristics as shown above. We nevertheless begin with cross-neighborhood comparisons because they provide a simple illustration of the main results and turn out to yield fairly similar estimates to those obtained below using our quasi-experimental design.

We compare earnings distributions across neighborhoods using our cross-sectional analysis sample, restricted to the years in which we have data on wage earnings from W-2's (1999-2009). We pool the observations for wage-earners across all years in this dataset and divide the ZIP-3-by-year cells into ten deciles based on sharp bunching rates, weighting by the number of observations in each cell. Figure 9 plots the distribution of W-2 wage earnings for individuals in the lowest and highest deciles of b_{ct}. Panel A considers EITC recipients with one child, while Panel B considers those with two or more children. The vertical lines denote the beginning and end of the refund-maximizing EITC plateau. In both panels, there is an increased concentration of the wage earnings distribution around the refund-maximizing region of the EITC schedule in areas in the top decile of sharp bunching b_{ct}. Under Assumption 2a, we can interpret this result as evidence that the EITC induces individuals to choose earnings levels that yield larger EITC refunds in high-knowledge areas.[49]

To characterize the excess mass more precisely, Figure 10 plots the difference between the earnings distributions for the highest and lowest deciles. For both the one child (Panel A) and 2+ child (Panel B) cases, the largest difference between the two densities occurs precisely in the refund-maximizing plateau region of the relevant schedule. As discussed above, audit studies reveal that W-2 earnings are rarely misreported, allowing us to interpret the differences in earnings distributions in Figure 10 as being driven by "real" labor supply choices rather than manipulation of reported income. The only potential source of misreporting on W-2's is for firms to collude with workers to misreport W-2 earnings to the IRS, for instance by paying workers part of their earnings off the books. While such collusion may be feasible in small family firms, it is much less likely to occur in large firms given the complexity of sustaining collusion on a large scale (Kleven, Kreiner, and Saez 2009). To ensure that our results are not driven by collusive reporting effects, we repeat the analysis in Figure 10 for wage-earners working in firms with 100 or more employees.

---

[49] One may be concerned that the behavioral response occurs through differences in child claiming behavior across areas rather than earnings behavior. For instance, if divorced couples in high-knowledge areas are more likely to claim a child on the return that produces a larger EITC refund, we would see differences in earnings distributions as in Figure 9. We address this source of selection bias in the next subsection by exploiting exogenous information on the date of child birth.

Within this subgroup, the difference in earnings distributions between the highest and lowest sharp-bunching areas is very similar to that in the full sample. We therefore conclude that the wage earnings changes in high-bunching areas are not driven by reported earnings manipulation.[50]

The analysis in Figures 9 and 10 considers only the first and tenth deciles of b_{ct}, the areas with the least and most knowledge about the EITC schedule. In Figure 11a, we extend the analysis to include all neighborhoods by plotting average EITC amounts for wage-earners vs. the level of sharp bunching b_{ct} in their ZIP-3-by-year cell. The average EITC refund effectively measures the concentration of the earnings distribution around the refund-maximizing region of the schedule.[51] Consistent with the earlier results, wage-earners in areas with high sharp bunching have earnings that produce significantly larger EITC refund amounts. A one percentage point increase in b_{ct} raises the EITC refund by $15.9 on average. Wage-earners in the highest-bunching areas earn EITC refunds that are on average $122 (5.1%) higher than those living in the lowest-bunching (near-zero knowledge) neighborhoods. Under Assumption 2a, this implies that behavioral responses to the EITC schedule raise EITC refund amounts by 5.1% in the highest bunching decile.

Cross-Neighborhood Movers. A natural approach to evaluate Assumption 2a and assess whether the level of knowledge in a neighborhood has a causal impact on earnings behavior is to again analyze changes in behavior for individuals who move across neighborhoods. Figure 11b plots changes in EITC refunds from the year before the move (event year -1) to the year after the move (event year 0) against the change in sharp bunching b_{ct} from the old to the new neighborhood. This figure exactly replicates Figure 6, restricting the sample to wage earners. Note that Figure 11b can be interpreted a first-differenced version of Figure 11a, relating changes in EITC refunds to changes in local knowledge for movers using our movers analysis sample.

Figure 11b shows that wage-earners who move to higher b_{ct} ZIP-3s change their earnings behavior so that their EITC refunds rise sharply. That is, increases in information in one's neighborhood lead to earnings responses that raise EITC refund amounts. In contrast, for individuals who move to areas with lower levels of sharp bunching, the slope of the relationship has, if anything, the oppo-

---

[50]One may be concerned that individuals in high-knowledge areas work in the formal sector up the point where they maximize their EITC refund and then work in informal jobs. Two pieces of evidence suggest that this is unlikely. First, our analysis of audit data (Chetty et al. 2012) shows that the likelihood of misreporting total earnings is no higher for individuals who report wage earnings in the plateau. Second, as we show below, most of the excess mass in the plateau comes from individuals raising W-2 earnings in the phase-in region in high-knowledge areas. The phase-in response cannot be driven by under-reporting of income from other jobs.

[51]EITC refund amounts also vary with marital status and number of children. Although differences in these demographics across areas could in principle affect the estimate in Figure 11a, we find very similar results within each of these demographic groups.

site sign.[52] We reject the null hypothesis that there is no kink in the slope of the control functions at 0 with $p < 0.0001$. This finding echoes the pattern of learning and memory documented above for the self-employed in Figure 6. The asymmetric persistence of past neighborhoods rules out a broad class of omitted variable biases that may arise from simple differences in characteristics across areas with different levels of sharp bunching. The finding that wage-earners making real decisions exhibit asymmetry also provides further evidence that the spatial heterogeneity in EITC response is due to knowledge about the schedule rather than tax compliance rates or other factors.[53]

While these findings show that neighborhoods have a causal effect on individuals' earnings behavior, they do not identify the extent to which individuals actively change their own behavior when exposed to more information about the EITC. Part of the increase in EITC refund amounts when individuals move to areas with higher b_{ct} in Figure 11b could in principle arise simply because individuals draw wage offers from a distribution that is more concentrated around the EITC plateau even if they do not actively reoptimize in response to the program incentives themselves.[54] We now turn to a research design that allows us to isolate individuals' responses to changes in incentives more precisely.

V.B   Impacts of Child Birth on Wage Earnings

In this section, we implement a second research design to characterize the impact of the EITC on wage earnings behavior that does not rely purely on cross-neighborhood comparisons.   Our strategy relies on the fact that individuals without children are eligible for only a very small EITC (see Section III) and therefore serve as a control group that can be used to net out differences across areas.   We implement this strategy by studying changes in earnings around the birth of a first child.   The first birth changes low-income families' incentives to earn significantly and is thus a powerful instrument for tax incentives.   The obvious challenge in using child birth as an

---

[52]The only parameter that is non-parametrically identified in this figure is the kink at 0.   The negative slope of the control function to the left of zero could be due to various factors that covary smoothly with the change in b_{ct}. For instance, because individuals who experience large drops in b_{ct} come from high bunching areas, differences across areas in movers' characteristics could generate differences in slopes.   The identifying assumption underlying inference from the kink is that any such correlated factors have smooth impacts on the slopes.

[53]For instance, one may be concerned that norms about tax compliance could have asymmetric persistence: once one observes someone else misreport earnings, it becomes an acceptable habit.   The asymmetric persistence of wage earnings rules out such models and implies that individuals' perception of incentives changes when they move to areas with high sharp bunching.

[54]Another potential concern is reverse causality: areas with wage earnings distributions that have substantial mass around the plateau for exogenous reasons may end up having higher sharp bunching as individuals near the plateau learn how to earn larger refunds.   It is difficult to explain the asymmetric pattern in Figure 11b purely with reverse causality, but it is possible that the magnitudes of the estimates obtained from cross-neighborhood comparisons are biased by such factors.

instrument for tax incentives is that it affects labor supply directly. We isolate the impacts of tax incentives by again using differences in knowledge across neighborhoods. In particular, we compare changes in earnings behavior around child birth for individuals living in areas with high levels of sharp bunching with those living in low-bunching areas. Low-bunching areas provide a counterfactual for how earnings behavior would change around child birth in the absence of the tax incentives.

We divide our child birth analysis sample into deciles based on sharp bunching in the individual's ZIP-3 in the year of child birth, as described in Section IV.D. Figure 12 plots W-2 wage earnings distributions for wage-earners in the year before (Panel A) and the year of first child birth (Panel B). The distributions are reported for those living in deciles 1, 5, and 10 of the sharp bunching distribution when they have a child. In the year before child birth, the wage earnings distributions are virtually identical across areas with low vs. high levels of sharp bunching.[55] However, an excess mass of wage-earners emerges around the plateau in high bunching areas immediately after birth, showing that individuals in these areas make an effort to obtain a larger EITC refund when making labor supply choices after child birth. Connecting this result to the cross-sectional correlations in Table II, Figure 12 essentially shows that individuals who live in areas with a high density of EITC filers have heard more about the credit by the time they have a child and therefore respond more strongly to its incentives.

The identification assumption underlying the research design in Figure 12 is that the direct impacts of child birth on earnings do not vary across neighborhoods with different levels of knowledge about the tax code (Assumption 2b). We assess the validity of this "common trends" assumption by examining trends prior to child birth using an event study design. Let year 0 denote the year in which the child is born (and hence the family becomes eligible for a larger EITC) and define event time relative to this year. Define an individual's simulated EITC credit as the EITC an individual would receive given her wage earnings if she had one child and were single. This simulated EITC credit is a simple statistic for the concentration of the wage earnings distribution around the EITC plateau.[56]

---

[55]In Table II, we showed that areas with higher sharp bunching have a higher density of EITC tax filers. This is not inconsistent with the result in Figure 12a. Figure 12a shows that the conditional earnings distributions among individuals just about to give birth are very similar across areas. However, the unconditional distributions differ across areas (e.g., because of differences in age and number of children). This is why we use an event study around child birth rather than comparisons of earnings distributions across all individuals with and without children for identification.

[56]We use the simulated credit with fixed parameters in this analysis rather than the actual credit to separate changes in earnings from mechanical changes in credit amounts when individuals have children.

Figure 13 plots the simulated EITC by event year for wage earners with incomes in the EITC-eligible range for exactly the same three groups as in Figure 12. For scaling purposes, we normalize the level of each series at the mean simulated credit in $t = -4$ by subtracting the decile-specific mean in $t = -4$ and adding back the mean simulated EITC across the three deciles in $t = -4$ to all observations. Simulated EITC amounts trend similarly in low, middle, and high bunching areas prior to child birth, supporting Assumption 2b. In the year of child birth, the simulated credit jumps significantly in high bunching areas relative to low bunching areas, showing that individuals in high-knowledge areas make an active effort to maintain earnings closer to the refund-maximizing level after having a child.[57] We estimate the magnitude of the impact using difference-in-differences specifications analogous to those used in the movers event studies in Figure 4a, clustering standard errors at the ZIP-3-by-birth-year level. EITC refunds increase by \$85.4 (4.7%) more from the year before to the year of child birth in the highest bunching decile relative to the lowest bunching decile.

In Figure 14a, we expand the analysis to include all neighborhoods by plotting the change in the simulated EITC from the year before birth (event year -1) to the year of birth (event year 0) vs. the level of sharp bunching in the individual's ZIP-3 in the year of birth, which we denote by b_{ct_0}. In this figure, we include all wage earners with incomes in the EITC-eligible range, as in Figure 12, as well as those with zero earnings (whose simulated credit is zero) to incorporate extensive margin responses. Consistent with the preceding evidence, individuals living in areas with higher b_{ct_0} (i.e., higher knowledge) have significantly larger increases in simulated EITC amounts around child birth. A one percentage point (0.58 standard deviation) increase in b_{ct_0} leads to a \$26.5 increase in the EITC after child birth, an effect that is statistically significant with $p < 0.0001$ with standard errors clustered at the ZIP-3-by-birth-year level.

Endogenous Sample Selection. Our child birth analysis sample makes two restrictions that could potentially lead to selection bias, thereby violating Assumption 2b. The first restriction is that we can only link parents to children they claim as dependents.[58] Because the decision to claim a child could be endogenous to knowledge about the EITC, this could also potentially bias our estimates through two channels. First, if a child is never claimed by any parent as a dependent, he or she is not included in our sample. In practice, over 97% of children are claimed as dependents on

---

[57]The slight divergence between the series in year -1 may occur because individuals in high-bunching areas keep their jobs prior to birth, recognizing that they will soon be eligible for a large EITC refund.

[58]Importantly, we observe date of birth from social security records. Each child's birth date is therefore measured independently of parents' tax filing behavior; only our link between parents and children is potentially endogenous to EITC incentives.

a tax return within 4 years of their birth.[59] Hence, endogeneity arising from whether a child is claimed at all is minimal.[60] Second, selection bias could arise if the person who claims a child is endogenously selected, e.g. if the family member who gets the highest EITC refund claims the child in high-knowledge areas. Such selection bias should be manifested in the period prior to child birth, as it would produce differences in simulated EITC credit amounts in event year -1 in Figures 12a and 13. Stated differently, we find sharp *changes* in earnings behavior within individuals around child birth. Bias can arise only if the decision to claim a child is related to changes in earnings around the time of child birth differentially across areas. While we cannot directly rule out such dynamic selection patterns, they are unlikely to produce a sharp break in earnings behavior only in the year of child birth given the smooth and relatively parallel dynamics of earnings across areas in prior years. Moreover, selection biases are unlikely to explain the asymmetric impacts of past neighborhoods for movers reported above.

The second restriction we impose above is to exclude individuals who report non-zero self-employment income in order to isolate wage earnings responses. If the choice to report self-employment income varies endogenously across areas, this restriction could also bias our estimates of the impact of the EITC on wage earnings.[61] To address this concern, we analyze changes in W-2 earnings around child birth for the full sample, including both wage-earners and the self-employed. We calculate the simulated EITC credit based purely on W-2 wage earnings even if the individual has self-employment income to isolate wage earnings responses. Figure 14b shows that the relationship between sharp bunching and the change in EITC amounts around child birth remains highly significant when the self-employed are included, with a point estimate of $19.4.[62] We use this technique to adjust for potential endogenous selection by including self-employed individuals and computing EITC amounts based on W-2 earnings in all the remaining

---

[59]We compute this statistic by comparing the total number of dependents claimed in the tax data to total births in the U.S. from vital statistics. This ratio is approximately 99% for births between 2000 to 2005. This figure slightly overstates our true coverage rate because it ignores children who immigrated to the U.S. and are claimed by their parents. Comparing vital birth statistics to all individuals recorded in the tax data, we estimate that immigration at young ages adds less than 0.5% per year to the size of a cohort, and hence obtain a lower bound of 97% for the fraction of individuals claimed.

[60]Most children are claimed very quickly after child birth presumably because knowledge that claiming children yields large tax credits is widespread. Conditional on claiming a child within four years of his or her birth, we find no evidence that parents living in ZIP-3's with high levels of sharp bunching b_{ct} claim a child more quickly after birth.

[61]For instance, suppose individuals in high-bunching areas are more likely to fabricate self-employment income after child birth if their wage earnings put them in the phase-in region rather than the plateau. By excluding those with self-employment income, we would artifically obtain a sample that exhibits more mass in the wage earnings distribution around the plateau in high-bunching areas.

[62]The magnitude of the coefficient is attenuated because we effectively miscalculate EITC amounts for self-employed individuals by using only their wage earnings to simulate their credits. We discuss how this attenuation bias can be corrected when computing elasticities below.

tables and figures.

Robustness Checks. In Table III, we assess the robustness of the result in Figure 14b to alternative specifications of the form:

$$(5) \qquad EITC_{ict} = \alpha + \beta_1 b_{ct_0} + \beta_2 \text{post} + \beta_3 \text{post} \times b_{ct_0} + \gamma X_{ict} + \varepsilon_{ict}$$

We estimate (5) using only observations in the year before and the year of child birth, $t \in \{-1, 0\}$. In this equation, $EITC_{ict}$ denotes the simulated credit individual i in ZIP-3 c obtains in event year t, post denotes an indicator for the year of child birth ($t = 0$), and X_{ict} denotes a vector of covariates. The coefficient of interest, β_3, measures the impact of a 1 percentage point increase in sharp bunching b_{ct_0} on the change in the simulated credit from the year before to the year after birth. Standard errors are clustered at the ZIP-3-by-birth-year level to account for potential correlation in earnings across residents of an area. Column 1 of Table III reports β_3 for the specification in Figure 14b (with no controls X_{ict}) as a reference.[63] Column 2 replicates column 1, restricting the sample to individuals working at firms with more than 100 employees (based on the number of W-2's). We continue to find a highly significant impact in this subgroup, confirming that these changes are not driven purely by manipulation of reported income. The magnitude of the effect is smaller because this specification excludes those with zero earnings from the sample, eliminating extensive margin responses. Column 3 adds ZIP-3 fixed effects interacted with the post indicator, so that β_3 is identified purely from variation in b_{ct_0} over time within areas.[64] The coefficient on b_{ct_0} remains large and highly significant in this specification, showing that unobservable differences across areas do not drive our findings.

A simple placebo test for our child birth research design is to examine changes in behavior for individuals having their third child instead of first child. Individuals with two or three children were eligible for the same EITC credit during the years of child birth that we analyze (2000 to 2005). The series in triangles in Figures 14a and 14b plots changes in simulated credit amounts (again using the one-child EITC schedule) from the year before to the year of the birth of a third child. Reassuringly, the relationship between neighborhood sharp bunching and changes in simulated credits around the birth of the third child is a precisely estimated zero, as shown in column 4 of Table III. This result confirms that the impacts of child birth on wage earnings do not

---

[63] In a balanced panel, the estimate of β_3 in equation (5) is identical to the estimate obtained from a univariate regression of the change in EITC amounts on b_{ct_0} as in Figure 14b.

[64] Allowing for ZIP-3×post fixed effects permits every ZIP-3 to have a different trend in $EITC$ around child birth. Hence, the only remaining source of identification for β_3 comes from comparing individuals who give birth in different years within a ZIP-3.

vary systematically across neighborhoods in the absence of changes in EITC incentives, supporting Assumption 2b.

The estimated impact from the child birth design of a \$19.4 increase in the simulated credit per percentage point of sharp bunching is similar to the corresponding cross-sectional estimate in Figure 11a of \$15.9. As discussed in Section II, cross-neighborhood comparisons incorporate endogenous changes in wage rates offered by firms as a result of shifts in the labor supply curve induced by the EITC. In contrast, changes in labor supply around child birth do not affect the equilibrium wage rate a new parent is offered as long as labor markets for parents and non-parents are not segregated. The fact that both research designs uncover significant and relatively similar impacts of the EITC on earnings suggests that general equilibrium feedback effects do not fully undo the partial-equilibrium changes in earnings behavior induced by the EITC. However, we cannot definitively identify the magnitude of general equilibrium effects because our cross-sectional estimate relies on a strong assumption for identification, namely that low and high bunching areas would have comparable wage earnings distributions absent the EITC (Assumption 2a).

Decomposition of Phase-In, Phase-Out, and Extensive Margin Responses. The welfare consequences of the EITC depend on whether the higher concentration of earnings around the refund-maximizing plateau of the EITC schedule comes from increased earnings for those who would have been in the phase-in region or reduced earnings from those who would have been in the phase-out region. To isolate the phase-in response, we define a "simulated phase-in credit" as the phase-in portion of the EITC schedule (for a single earner with one child) combined with a constant refund above the first kink at the refund-maximizing level. Analogously, we define a "simulated phase-out credit" as the phase-out portion of the schedule combined with a constant refund below the second kink at the refund-maximizing level. Appendix Figure 4 depicts these two schedules.[65] The simulated phase-in credit is a convenient summary statistic for earnings increases in the phase-in region because it grows when individuals raise their earnings in the phase-in but is unaffected by changes in earnings in the plateau and phase-out regions. The simulated phase-in credit asks, "How would behavioral responses affect refund amounts if the EITC stayed constant at its maximum level and was never phased out?" The simulated phase-out credit similarly isolates changes in earnings behavior in the phase-out region. We define both simulated credits based purely on wage earnings (but include self-employed individuals in the sample) as above.

---

[65]Formally, we define the simulated phase-in credit as $\min(.34 \times z_i, 3050)$ and the phase-out credit as $\max(3050 - .16 \times \min(z_i - 16450, 0), 0)$.

Figure 15a plots changes in the simulated phase-in and phase-out credits around child birth vs. the degree of sharp bunching. The corresponding regression coefficients are reported in columns 5 and 6 of Table III. By construction, the slopes of the two coefficients sum to the slope for the full EITC credit schedule in column 1. Figure 15a shows that \$14.2/\$19.4 = 73% of the increase in EITC refunds in high-bunching areas comes from the phase-in region. As a result, the EITC program is successful in increasing wage earnings for low-income individuals despite creating high marginal tax rates in a broad part of income distribution.

Next, we separate intensive and extensive margin responses. To analyze extensive margin responses, we define "working" as having positive W-2 earnings in a given year. We use the full sample (including non-workers, self-employed individuals, and wage earners) for this analysis. Figure 15b plots the change in the fraction of individuals working from the year before to the year of child birth vs. sharp bunching; the corresponding regression coefficient is reported in column 7 of Table III.[66] Consistent with prior studies, we find significant extensive margin responses. Individuals living in areas with high levels of sharp bunching are more likely to continue working after they have a child than those living in areas with little sharp bunching. To gauge the extent to which extensive margin responses contribute to the increase in EITC refunds, we assume that extensive margin entrants earn the average EITC refund in the child birth sample conditional on working (\$1,075). Under this assumption, the extensive margin contributes \$5.8/\$19.4 = 29% to the increase in EITC refunds, as shown in column 8 of Table III.[67]

Finally, in column 9 of Table III, we analyze the number of W-2's per individual, which is a proxy for the number of distinct jobs an individual held over the year. A one percentage point increase in sharp bunching leads to a 0.017 (0.018 SD) increase in the number of W-2's filed after child birth. Hence, part of the increase in earnings in the phase-in region comes from individuals taking additional part-time jobs.[68] Adjustment in part-time jobs could explain why earnings responses to the EITC are larger in the phase-in than the phase-out. In our child birth sample, individuals in the phase-in have 1.61 W-2's per person, at which they earn \$2,300 per job on average. Those

---

[66] The mean fraction of individuals working in this sample is 82% in the year before child birth and 84% in the year of child birth. The fraction working increases around child birth because this sample includes predominantly young, unmarried women who are entering the labor force and because our definition of "working" is defined as having *any* earnings over a year.

[67] We can obtain a non-parametric upper bound on the extensive margin response by assuming that all individuals who enter on the extensive margin earn the maximum EITC refund (i.e., choose a level of earnings in the plateau). This calculation reveals that the extensive margin response accounts for at most 90% of the total response. Hence, we can be confident that the EITC induces responses on both the intensive and extensive margin; however, the relative magnitude of intensive and extensive responses is less clear.

[68] It is difficult to determine exactly what fraction of the response comes from additional jobs because we would need an estimate of earnings at the marginal job.

in the phase-out have 1.42 W-2's with mean earnings of \$14,300 per W-2. Because they work more small, part-time jobs, individuals in the phase-in may be able to change their earnings more easily than those in the phase-out. An alternative explanation for larger phase-in elasticities is that current perceptions of the EITC schedule focus on phase-in incentives more than the phase-out incentives. We cannot distinguish between these explanations with our research design.

VI   Policy Calculations

In this section, we use our estimates to quantify the impacts of the EITC in two ways. First, we calculate the elasticities implied by our analysis in a neoclassical model to gauge the magnitudes of the earnings responses documented above. Second, we characterize the impacts of the EITC on the wage earnings distribution and poverty rates.

Earnings Elasticities. One of the main lessons of our study is that the impacts of tax policies cannot be characterized using a single elasticity, as the behavioral responses we have documented do not conform to the predictions of traditional labor supply models. Nevertheless, to help gauge magnitudes and revenue consequences, we calculate the elasticity that would generate the increase in EITC refunds we observe under a neoclassical, frictionless model.[69]

Panel A of Table IV reports elasticity estimates for wage earners. The first column reports the intensive-margin elasticity that would generate an increase in EITC refunds commensurate to the empirical estimates above. We compute these elasticities as follows. With a standard iso-elastic labor supply function, a frictionless model with elasticity ε implies

$$\log(z + \Delta z) - \log(z) = \varepsilon \cdot \log(1 - \tau)$$

where τ is the actual marginal tax rate an individual faces because of the EITC, z is the level earnings when the EITC marginal tax rate is perceived to be zero, and $z + \Delta z$ is earnings when the EITC marginal tax rate is accurately perceived to be τ.[70] The change in the EITC refund

---

[69]Our technique is analogous to that used by Ashenfelter (1983) to estimate the elasticities implied by behavioral responses to the Negative Income Tax (NIT). Ashenfelter calculates the elasticities that would generate the observed changes in NIT participation rates under a constant-elasticity labor supply model; here we calculate the elasticities that would generate the observed increase in EITC refunds in the same model. Note that in a frictionless labor supply model, the increase in EITC refunds would come primarily from a point mass in the wage earnings distribution at the kink points of the EITC schedule, which is not what we observe empirically. This is why our estimate does not represent the actual structural labor supply elasticity implied by the data.

[70]For simplicity, this equation assumes that individuals remain on the interior of the budget segment when they increase earnings by Δz. Accounting for the kinks in the EITC schedule significantly complicates the calculations and has little impact on the estimated elasticities.

induced by the earnings response is:

$$-\tau \cdot \Delta z = -\tau \cdot [(1 - \tau)^{\varepsilon} - 1] \cdot z$$

and the mean increase in EITC refunds due to behavioral responses in the phase-in and phase-out regions is

(6) $$\Delta EITC = -\phi_1 \tau_1 \cdot [(1 - \tau_1)^{\varepsilon} - 1] \cdot z_1 - \phi_2 \tau_2 \cdot [(1 - \tau_2)^{\varepsilon} - 1] \cdot z_2$$

where $\phi_1 = 26.9\%$ and $\phi_2 = 22.1\%$ denote the fraction of individuals in the phase-in and phase-out regions in the year after birth in our child birth sample.[71] $\tau_1 = -34\%$ and $\tau_2 = 16\%$ denote the phase-in and phase-out marginal tax rates and $z_1 = \$5,725$ and $z_2 = \$23,216$ denote the mean earnings levels in the phase-in and phase-out regions.

We calculate the change in EITC refunds $\Delta EITC$ as follows. We begin from our estimate that a one percentage point increase in sharp bunching increases mean EITC amounts by \$19.4, shown in column 1 of Table III. This estimate uses the full sample, in which 10.8% of individuals are self-employed. As in our model, we assume that self-employed individuals do not respond along the wage earnings margin (as adjusting self-employment income is less costly). Therefore, the impact of the EITC on the treated (i.e., the wage earners) is $\$19.4/(1 - .108) = \$21.7$. The average ZIP-3 in our sample has a level of sharp bunching b_{ct} that is 1.34 percentage points higher than a bottom decile bunching area. Under our maintained assumption that areas in the bottom decile of bunching exhibit no behavioral response to the EITC, the increase in EITC refunds for the average neighborhood in the U.S. relative to areas with no response is $\Delta EITC = 1.34 \times 21.7 = \$29.1$. Substituting this value into (6) and solving for ε yields $\varepsilon = 0.10$ (Table IV, Column 1, Row 1).

This estimate of ε assumes that the earnings elasticity is the same in the phase-in and phase-out regions of the schedule. However, as demonstrated in Figure 15a, responses in the phase-in and phase-out regions are quite different in magnitude. In columns 2 and 3 of Table IV, we estimate the elasticities in the phase-in and phase-out regions separately using the estimates from columns

---

[71] In our child birth sample, 26.9+22.1=49% of individuals have income below the end of the phase-out region. For simplicity, we abstract from the constant marginal tax rate in the plateau region and assume that those in the bottom half of the plateau are in the phase-in and those in the upper half of the plateau are in the phase-out in terms of the change in the marginal incentives they face.

5 and 6 of Table III. We estimate these elasticities using the formulas

$$\Delta\text{Phase-in EITC} = -\phi_1 \cdot \tau_1 \cdot [(1-\tau_1)^{\varepsilon_1} - 1] \cdot z_1$$

$$\Delta\text{Phase-out EITC} = -\phi_2 \cdot \tau_2 \cdot [(1-\tau_2)^{\varepsilon_2} - 1] \cdot z_2$$

Computing the changes in EITC amounts as above, we obtain $\varepsilon_1 = 0.14$ for the phase-in elasticity and $\varepsilon_2 = 0.06$ for the phase-out elasticity.

Finally, in column 4 of Table IV, we report estimates of extensive margin elasticities. We define the participation tax rate τ_{ext} as the mean EITC refund as a percentage of mean income conditional on working. We then use the estimate in column 7 of Table III and define $\hat{\varepsilon}_{\text{ext}}$ as the log change in participation rates (starting from the sample mean) divided by the log change in the net-of-participation-tax rate. This yields an estimate of $\hat{\varepsilon}_{\text{ext}} = 0.10$. In interpreting this elasticity, it is important to note that our definition of "working" is having positive earnings at any time within a year. While this annual concept is what matters for optimal policy in an annual tax system, it is likely that the EITC induces larger extensive margin responses at higher frequencies, e.g. at the weekly level. This could explain why our estimate of the extensive margin elasticity is slightly smaller than estimates in prior studies using survey data (e.g., Eissa and Liebman 1996).

The preceding elasticities apply to the U.S. as a whole given the average level of knowledge about the EITC schedule in the economy between 2000 and 2005. In row 2 of Table IV, we report elasticities for areas in the top decile of sharp bunching b_{ct}, i.e. the areas with the highest levels of knowledge in our sample. We calculate these elasticities using the same method as above, but define the increase in EITC refunds as the difference-in-differences between EITC refunds in the top vs. bottom decile in the year after birth vs. before birth. The elasticities are roughly 5 times larger in areas in the top bunching decile relative to the country as whole. These calculations suggest that behavioral responses may grow significantly as knowledge about the EITC's structure continues to spread across the U.S., increasing earnings levels as well as expenditures on the program.

Panel B of Table IV replicates Panel A including self-employment income. It reports total earnings elasticities, calculated exactly as in Panel A using regressions analogous to those in columns 1 and 8 of Table III but using total earnings instead of wage earnings. These regression estimates are reported in Appendix Table I. Total earnings elasticities are much larger because self-employed individuals exhibit large responses to the EITC, especially in high bunching areas. The mean earnings elasticity is 0.22 in the U.S. as a whole and 0.95 in the top bunching decile. Even though less than a fifth of EITC claimants are self-employed, they account for a substantial fraction of the

increase in EITC refunds via behavioral responses. As shown in our companion paper (Chetty et al. 2012), much of this response is driven by non-compliance. Reducing the behavioral response due to non-compliance through auditing or changes in reporting requirements for self-employment income may make the EITC more effective at raising true earnings.

Impacts on the Income Distribution. We characterize the impact of the EITC on the wage earnings distribution by calculating the fraction of wage-earners in our cross-sectional analysis sample below the poverty line and other income thresholds $\bar{z}$.[72] We begin by estimating the causal impact of the EITC on the fraction of individuals below each threshold $F(\bar{z})$ using our child birth design. Let $t = 0$ denote the year of child birth and $t = -1$ the year before child birth. Let $b_d = 1$ denote ZIP-3-by-birth-year cells in the first decile of the bunching distribution. We define the treatment effect of the EITC using a difference-in-differences estimator:

$$(7) \qquad \Delta F(\bar{z}) = [F(\bar{z}, t = 0) - F(\bar{z}, t = -1)] - [F(\bar{z}, t = 0|b_d = 1) - F(\bar{z}, t = -1|b_d = 1)]$$

The first difference is the change in the fraction of individuals below the poverty line in the full population; the second is the same difference within neighborhoods in the lowest bunching decile. We estimate the fraction who would have wage earnings below threshold $\bar{z}$ absent the EITC in the full population as $F(\bar{z}) - \Delta F(\bar{z})$.

We characterize the impact of the EITC on the average earnings distribution between 2000 and 2005, the period over which we estimate the treatment effect $\Delta F(\bar{z})$ using our child birth sample. The first row of Table V shows our estimate of $F(\bar{z})$ without the EITC for various multiples of the poverty line. For instance, we estimate that 31.9% of wage-earners in our cross-sectional analysis sample – which consists of EITC-eligible households with children – would be below the poverty line without the EITC. In the second row, we add in EITC payments based on the individual's wage earnings, marital status, and number of dependents.[73] We assume that all eligible households claim their benefit and hold wage earnings for each household fixed at the same level as in the first row. The difference between the first and second rows thus reflects the mechanical effect of EITC payments on post-tax incomes. EITC payments shift the income distribution upward significantly; the fraction below the poverty line falls to 22.0%.[74] The third row reports statistics

---

[72] We use the official poverty line in each year from 2000-2005 (to match the years of the birth sample) corresponding to the individual's marital status and number of children.

[73] When making this calculation, we assume that the treatment effect $\Delta F(\bar{z})$ is constant in percentage terms across all subgroups.

[74] As 15% of all households with children in the U.S. are EITC eligible, the EITC reduces overall poverty rates in the population by approximately 2 percentage points.

for the observed post-EITC income distribution in the aggregate economy. This row incorporates behavioral responses to the EITC on top of the mechanical effects in the second row. Behavioral responses to the EITC further increase incomes at the lowest levels, as workers response to the marginal subsidy on the phase-in. Taking behavioral responses into account, the fraction below the poverty line with the EITC is 21.0%.

In the last row of Table V, we consider the effect of increasing knowledge of the EITC everywhere to the level observed in the highest sharp bunching decile. This row asks, "How would the EITC affect the earnings distribution in the U.S. if knowledge about the schedule were at the level in the highest bunching decile?" We estimate this effect by recalculating (7), replacing the first term with the CDFs in the top bunching decile instead of the full sample. We then add this causal effect back to the counterfactual distribution calculated in the first row of Table V and recompute EITC refund amounts. The increased level of knowledge triples the behavioral response to the EITC, further lowering the fraction below the poverty line to 19.6%.

Table V yields three main lessons. First, the impacts of the EITC on inequality come largely through its mechanical effects rather than behavioral responses in the nation as a whole. Second, behavioral responses tend to reinforce the mechanical effects of the EITC in raising incomes of the lowest earning households in the U.S. For instance, the fraction earning less than half the poverty line – which is near the end of the phase-in region – falls from 13.7% to 9.4% due to the mechanical transfer and falls further to 8.2% because individuals in the phase-in raise their earnings. In contrast, behavioral responses to the disincentive effects of the EITC in the phase-out region of the schedule have much smaller impacts: the fraction earning less than 200% of the poverty line falls from 77.3% to 71.1% due to the mechanical effect, but rises to only 71.3% when incorporating behavioral responses. Third, more than a decade after the EITC was implemented in its current form, the aggregate response to the EITC still comes from a relatively small subset of neighborhoods in the U.S. in which behavioral responses are quite large. However, knowledge about the EITC – as measured by the level of sharp bunching – is still rising sharply, as shown in Figure 7a. As knowledge about the EITC continues to spread through the U.S., the EITC is likely to have larger effects on the aggregate income distribution.

VII   Conclusion

A growing literature finds that many policies have diffuse effects on economic behavior that are inconsistent with neoclassical models because of inattention and other frictions. Identifying diffuse

impacts has thus emerged as one of the major challenges for applied work on policy evaluation. This paper has developed a new method of addressing this challenge by using differences across neighborhoods in knowledge about the policy to obtain counterfactuals for diffuse responses. We apply this method to characterize the impacts of the EITC on earnings behavior by using the degree of sharp bunching at the refund-maximizing income level by the self-employed as a proxy for local knowledge about the EITC schedule. We find that areas with higher levels of knowledge exhibit significantly more mass in the wage earnings distribution around the EITC plateau. In addition, changes in marginal incentives due to child birth have larger impacts on wage earnings behavior in areas with higher levels of knowledge about the EITC.

The wage earnings response comes primarily from intensive-margin increases in earnings by individuals in the phase-in region. As a result, behavioral responses to the EITC reinforce its direct impacts in raising the incomes of low-income families with children. Overall, we conclude that the EITC has increased earnings and net income levels among low-income families in the U.S., with especially large impacts in areas with a high density of EITC claimants.

Our analysis can be extended and generalized in several dimensions. Most directly, one could use the counterfactuals developed here to study the impacts of the EITC on other behaviors, such as contribution to tax-deferred savings accounts, family formation, and earnings dynamics. One could also use a similar approach to develop proxies for knowledge about other policies and study their impacts. For instance, several studies have documented sharp bunching around the kinks of the Social Security earnings test schedule (Friedberg 2000, Gruber and Orszag 2003, Haider and Loughran 2008). Using spatial variation in such bunching, one may be able to characterize the impacts of Social Security incentives on retirement behavior in the U.S. Similar techniques could also shed light on the impacts of corporate tax credits, which create sharp incentives for manipulation around thresholds (e.g., Goolsbee 2004), but may affect real investment decisions more diffusely. More generally, using low-knowledge groups as counterfactuals could help uncover the impacts of a variety of important policies whose effects have proven difficult to characterize with traditional research designs.

References

Andreoni, James, Brian Erard, and Jonathon Feinstein (1998). "Tax Compliance," Journal Economic Literature, 36:2 : 818-860.

Ashenfelter, Orley (1983) "Determining Participation in Income-Tested Social Programs," Journal of the American Statistical Association , 78(383): 517-525.

Bertrand, Marianne, Erzo F. P. Luttmer, and Sendhil Mullainathan (2001) "Network Effects And Welfare Cultures," The Quarterly Journal of Economics 115(3), 1019-1055.

Bises, Bruno (1990). "Income Tax Perception and Labour Supply in a Sample of Industry Workers," Public Finance, 45(1): 3-17.

Bollinger, Christopher, Luis Gonzalez, and James P. Ziliak. 2009. "Welfare Reform and the Level and Composition of Income." In Welfare Reform and its Long-Term Consequences for America's Poor, James P. Ziliak, ed, Cambridge, UK: Cambridge University Press, 59-103.

Brown, C.V. (1968). "Misconceptions About Income Tax and Incentives," Scottish Journal of Political Economy, 16(2): 1-12.

Card, David and David Lee. "Regression Discontinuity Inference with Specification Error" Journal of Econometrics, 142(2): 655-674.

Chetty, Raj (2012). "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply." Econometrica 80(3): 969-1018.

Chetty, Raj, John N. Friedman, Peter Ganong, Alan H. Plumley, Kara E. Leibel, and Emmanuel Saez (2012), "Taxpayer Response to the EITC: Evidence from IRS National Research Program," Harvard University mimeo..

Chetty, Raj, John N. Friedman, Tore Olsen, and Luigi Pistaferri (2011). "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records," Quarterly Journal of Economics, 126(2): 749-804.

Chetty, Raj and Emmanuel Saez (2012), "Teaching the Tax Code: Earnings Responses to an Experiment with EITC Recipients," forthcoming, American Economic Journal: Applied Economics.

Cilke, James (1998). "A Profile of Non-Filers." U.S. Department of the Treasury, Office of Tax Analysis Working Paper No. 78.

Eissa, Nada and Hilary Hoynes (2004). "Taxes and the Labor Market Participation of Married Couples: The Earned Income Tax Credit," Journal of Public Economics, 88(9-10): 1931-1958.

Eissa, Nada and Hilary Hoynes (2006). "Behavioral Responses to Taxes: Lessons from the EITC and Labor Supply," In James Poterba, ed. Tax Policy and the Economy 20, Cambridge: MIT Press, pp. 74-110.

Eissa, Nada and Jeffrey Liebman (1996). "Labor Supply Response to the Earned Income Tax Credit," Quarterly Journal of Economics, 111(2): 605-637.

Friedberg, Leora (2000). "The Labor Supply Effects of the Social Security Earnings Test," The Review of Economics and Statistics, 82 (1): 48-63.

Fujii, Edwin T. and Clifford B. Hawley (1988). "On the Accuracy of Tax Perceptions," Review of Economics and Statistics, 70(2): 344-347.

Gelber, Alexander, and Joshua W. Mitchell (2012) "Taxes and Time Allocation: Evidence from Single Women and Men," Review of Economic Studies forthcoming.

Goolsbee, Austan (2004). "The Impact of the Corporate Income Tax: Evidence from State Organizational Form Data," Journal of Public Economics, 88(11): 2283-99.

Grogger, Jeffrey (2003). "The Effects of Time Limits, the EITC, and Other Policy Changes on Welfare Use, Work, and Income Among Female-Headed Families," Review of Economics and Statistics 85 (2), 394-408.

Gruber, Jonathan and Peter Orszag (2003). "Does the social security earnings test affect labor supply and benefits receipt?" National Tax Journal, 4(56):755-773.

Haider, Steven J. and David Loughran (2008). "The Effect of the Social Security Earnings Test on Male Labor Supply: New Evidence from Survey and Administrative Data." Journal of Human Resources 43(1):57-87.

Hausman, Jerry A. (1981). "Labor Supply," in How Taxes Affect Economic Behavioral, Henry Aaron and Joseph Pechman eds.Washington D.C.: Brookings Institute.

Hotz, V. Joseph, Charles H. Mullin, and John Karl Scholz (2011). "Examining the Effect of The Earned Income Tax Credit on the Labor Market Participation of Families on Welfare." Duke University Working Paper.

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

Hotz, V. Joseph and John Karl Scholz (2006). "Examining the Effect of the Earned Income Tax Credit on the Labor Market Participation of Families on Welfare." NBER Working Paper 11968.

Internal Revenue Service (1996). Federal Tax Compliance Research: Individual Income Tax Gap Estimates for 1985, 1988, and 1992, Publication 1415 (Rev. 4-96), Government Printing Press: Washington, D.C.

Internal Revenue Service (2011a). Statistics of Income: Individual Income Tax Returns, 2009 Publication 1304, Government Printing Press: Washington, D.C.

Internal Revenue Service (2011b). Statistics of Income: Individual Income Tax Returns, 2009 Publication 596, Government Printing Press: Washington, D.C.

Kleven, Henrik, Claus T. Kreiner, and Emmanuel Saez (2009). "Why Can Modern Governments Tax So Much? An Agency Model of Firms as Fiscal Intermediaries," NBER Working Paper 15218.

Kleven, Henrik J., and Mazhar Waseem (2012). "Behavioral Responses to Notches: Evidence from Pakistani Tax Records," London School of Economics Working Paper, April 2012.

Kopczuk, Wojciech, and Christian Pop-Eleches (2007). "Electronic Filing, Tax Preparers and Participation in the Earned Income Tax Credit," Journal of Public Economics. 91:1351-1367.

Leigh, Andrew (2010). "Who Benefits from the Earned Income Tax Credit? Incidence among Recipients, Coworkers and Firms," The B.E. Journal of Economic Analysis & Policy, Berkeley Electronic Press, 10(1).

Liebman, Jeffrey (1998). "The Impact of the Earned Income Tax Credit on Incentives and the Income Distribution," in James Poterba, ed. Tax Policy and the Economy 12, Cambridge: MIT Press, pp. 83-123.

Liebman, Jeffrey and Richard Zeckhauser (2004). "Schmeduling", Harvard University mimeo.

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

Meyer, Bruce (1995). "Natural and Quasi- Experiments in Economics." Journal of Business and Economic Statistics. 13(2): 151-161.

Meyer, Bruce (2010). "The Effects of the Earned Income Tax Credit and Recent Reforms." Ed. Jeffrey Brown, Tax Policy and the Economy, 24(1), Cambridge: MIT Press, 153-180.

Meyer, Bruce and Dan Rosenbaum (1999). " Welfare, the Earned Income Tax Credit, and the Labor Supply of Single Mothers." NBER Working Paper 7363.

Meyer, Bruce and Dan Rosenbaum (2001). "Welfare, the Earned Income Tax Credit, and the Labor Supply of Single Mothers." Quarterly Journal of Economics, 116(3): 1063-1114.

Plueger Dean (2009). "Earned Income Tax Credit Participation Rate for Tax Year 2005," Internal Revenue Service, available at http://www.irs.gov/pub/irs-soi/09resconeitcpart.pdf.

Ross Phillips, Katherin (2001). "Who Knows About the Earned Income Tax Credit?" Urban Institute policy brief, No. B-27, January.

Rothstein, Jesse (2010). "Is the EITC as Good as an NIT? Conditional Cash Transfers and Tax Incidence," American Economic Journal: Economic Policy, 2(1), 177–208.

Saez, Emmanuel (2010). "Do Taxpayers Bunch at Kink Points?" American Economic Journal: Economic Policy, 2(3), 180–212.

Saez, Emmanuel, Joel B. Slemrod, and Seth H. Giertz (2012). "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review," Journal of Economic Literature, 50(1), 3-50.

Slemrod, Joel (2007): "Cheating Ourselves: The Economics of Tax Evasion," Journal of Economic Perspectives, 21, 25–48.

Smeeding, Timothy M., Katherin Ross Phillips, and Michael A. O'Connor (2002). "The Earned Income Tax Credit: Expectation, Knowledge, Use, and Economic and Social Mobility," in Bruce Meyer and Douglas Holtz-Eakin, eds. Making Work Pay, Russell Sage Foundation: New York, also in National Tax Journal 53(4): 1187–1209.

Stephens-Davidowitz, Seth. "The Effects of Racial Animus on Voting: Evidence Using Google Search Data." Harvard University mimeo, 2011.

U.S. Census Bureau (2012) "Internet Use in the United States: October 2009." http://www.census.gov/hhes/computer/publications/2009.html

TABLE I

Summary Statistics for Cross-Sectional Analysis Sample, 1999-2009

Variable	Mean (1)	Std. Dev. (2)
Income Measures		
Total Earnings	$20,091	$10,784
Wage Earnings	$18,308	$12,537
Self-Employment Income	$1,770	$6,074
Indicator for Non-Zero Self-Emp. Income	19.6%	39.7%
Number of W-2's	1.32	0.94
Tax Credits		
EITC Refund Amount	$2,543	$1,454
Claimed EITC	88.9%	31.4%
Tax Professional Usage	69.6%	46.0%
Demographics		
Age	37.3	13.3
Number of Children	1.7	0.8
Married	30.3%	45.9%
Female (for single filers)	73.0%	44.4%
Neighborhood (ZIP-3) Characteristics		
Self-Emp. Sharp Bunching	2.05%	1.73%
EITC Filer Density	0.22	0.61
State EITC Top-Up Rate	5.00%	9.17%
Number of Observations	219,742,011	

Notes: This table reports summary statistics for the cross-sectional sample, which includes primary filers in our core sample (defined in Section 3) who file a tax return, report one or more children, and have income in the EITC-eligible range. We restrict the sample to 1999-2009, the years for which we have W-2 earnings data. Total earnings, which includes wage earnings and self-employment earnings, is the earnings measure used to calculate EITC refunds. Self-employment income is income reported on Schedule C. Wage earnings are earnings reported on Form W-2 by employers. We trim all income measures at -$20K and $50K. Tax professional usage is the fraction of individuals using a third-party tax preparer. Age is defined as of December 31 of a given tax year. Number of children is number of EITC-eligible dependents claimed on Schedule EIC; for those who do not file Schedule EIC, it is the number of non-elderly dependents claimed on Form 1040. Statistics for neighborhood variables weight ZIP-3 level means by the number of EITC-eligible individuals with children in the cross-sectional analysis sample. Self-employed sharp bunching is the fraction of EITC-eligible filers with children who both report total earnings within $500 of the first kink point in the EITC schedule and have non-zero self-employment earnings. EITC filer density is the number of EITC-eligible filers (measured in 1000's) per square mile in tax year 2000. State EITC top-up rate is state EITC as a fraction of the federal credit.

TABLE II
Cross-Sectional Correlates of Sharp Bunching

Dependent Variable:	Self-Employed Sharp Bunching Rate in ZIP-3 (%)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
EITC Filer Density in ZIP-3	1.93 (0.05)	1.82 (0.05)			0.44 (0.06)	0.69 (0.06)		
Tax Professional Usage in ZIP-3			9.86 (1.47)		3.02 (0.51)	3.46 (0.56)		
Google Search Intensity for "Tax"				0.30 (0.05)	0.14 (0.03)	0.19 (0.03)		
State EITC Top-Up Rate							0.07 (0.05)	
State Non-Compliance Rate								-1.51 (5.32)
Demographic Controls		x			x	x		
State Fixed Effects						x		
Year	2000	2000	2008	2008	2008	2008	2000	2000
R-squared	0.603	0.798	0.169	0.032	0.728	0.848	0.105	0.002
Number of ZIP-3s	873	873	883	875	870	849	886	51

Notes: Each column reports estimates from an OLS regression run at the ZIP-3 level, weighted by the number of individuals in each ZIP-3 in the cross-sectional analysis sample. Standard errors are reported in parentheses. EITC filer density is the number of EITC filers (measured in 1000's) per square mile in the ZIP-3. Tax professional usage is the fraction of EITC filers who use a professional tax preparer in the ZIP-3. Google search intensity for "tax" is the fraction of all Google searches in the ZIP-3 for phrases that include the word "tax" divided by standard deviation of this measure, so that the variable is scaled in standard deviation units. State EITC top-up rate is the size of the state EITC top-up as a fraction of the federal EITC; states without a state EITC are coded as zero. State non-compliance rate is the fraction of non-EITC-eligible individuals in a state with a difference between reported and corrected income greater than $1,000; this variable is measured using data from the 2001 IRS National Research Program audit data. The specification in column 8 is estimated at the state level because the non-compliance variable is only available by state even though it may vary locally. Note that state EITC top-up is also measured at the state level, but since that variable does not vary within state, we run the regression at the ZIP-3 level and cluster standard errors by state. The demographic controls include the percentage of the population that is foreign-born, white, black, Hispanic, Asian, and other. We use data from year 2000 in some specifications because Census data are available only in 2000; we use data from year 2008 in other specifications because Google search intensity was high only in more recent years.

TABLE III

Impacts of EITC on Wage Earnings: Regression Estimates from Child Birth Design

Dependent Variable:	Baseline Specification	Large Firms Only	With ZIP-3 Fixed Effects	Placebo Test: 3rd Child	Phase-in vs. Phase-out		Extensive Margin		
	Simulated EITC Refund				Simulated Phase-in Credit	Simulated Phase-out Credit	Positive W-2 Earnings	Mean EITC x (W-2 Earn > 0)	Number of W-2 Forms
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ZIP-3 Self-Emp. Sharp Bunching (%)	$19.4 (1.61)	$14.4 (1.14)	$34.7 (3.20)	-$1.89 (0.63)	$14.2 (1.55)	$5.2 (0.69)	0.54% (0.05)	$5.81 (0.52)	0.017 (0.002)
ZIP-3 by Post-Birth Fixed Effects			x						
Observations (millions)	29.96	13.20	29.96	10.07	29.96	29.96	29.96	29.96	29.96
Mean Level of Dep. Var. in Year Before Birth	$1,038	$1,209	$1,038	$899	$1,038	$1,038	84.8%	$960	1.78

Notes: All specifications are estimated using the child birth sample, which includes individuals in the core sample who had their first child between 2000 and 2005, using only the year before and the year of child birth. All columns include all individuals (wage earners, self-employed, and non-workers). Each column reports estimates from an OLS regression of an outcome on the level of sharp bunching in the ZIP-3-by-year cell in which the individual gives birth to his or her first child, an indicator for the post-birth year, and an interaction of sharp bunching and the indicator for the post-birth year. The table reports coefficients on the interaction term, which can be interpreted as the impact of a one percentage point increase in sharp bunching on the change in the outcome around child birth. Standard errors, clustered at the ZIP-3-by-birth-year level, are reported in parentheses. In column 1, the dependent variable is the simulated EITC refund. To calculate the simulated EITC refund, we apply the one-child EITC schedule for single filers to total household W-2 earnings, regardless of the household's actual structure and self-employment income. Column 2 replicates column 1, restricting the sample to individuals whose W-2 forms are all issued by firms with 100 or more employees in a given year. Column 3 adds ZIP-3 fixed effects to the specification in column 1. Column 4 replicates column 1 using individuals having 3rd births instead of 1st births (for whom there is no change in EITC in tax years 2000-2005) as a placebo test. The dependent variable in column 4 is again the one-child simulated EITC. Columns 5 and 6 decompose the response into the phase-in and phase-out regions. In column 5, the dependent variable is the simulated phase-in credit, which is calculated based on W-2 earnings using the schedule shown in Appendix Figure 4a. In column 6, the dependent variable is the simulated phase-out credit, calculated based on W-2 earnings using the schedule shown in Appendix Figure 4b. The estimates in columns 5 and 6 mechanically sum to the estimate reported in column 1. The dependent variable in column 7 is an indicator for having positive W-2 wage earnings. The dependent variable in column 8 is this indicator multiplied by the average EITC amount for wage earners conditional on working, which is $1,075 in this sample. The estimate in this column can be used to calculate the fraction of the response in column 1 that is due to extensive margin responses. The dependent variable in column 9 is the number of W-2 forms of the individual parent (not the tax return). The bottom row displays the average level of the dependent variable in the year before birth.

TABLE IV

Elasticity Estimates Based on Change in EITC Refunds Around Birth of First Child

	Mean Elasticity	Phase-in Elasticity	Phase-out Elasticity	Extensive Elasticity
	(1)	(2)	(3)	(4)
A. Wage Earnings				
Elasticity in U.S. 2000-2005	0.10 (0.008)	0.14 (0.011)	0.06 (0.006)	0.10 (0.009)
Elasticity in top decile ZIP-3s	0.46 (0.017)	0.58 (0.021)	0.30 (0.021)	0.59 (0.033)
B. Total Earnings				
Elasticity in U.S. 2000-2005	0.22 (0.013)	0.34 (0.020)	0.08 (0.004)	0.18 (0.012)
Elasticity in top decile ZIP-3s	0.95 (0.026)	1.32 (0.036)	0.34 (0.012)	1.05 (0.039)

Notes: The first panel reports elasticities using wage earnings responses estimated in Table III; the second panel reports elasticities using total earnings responses (including self-employment income) estimated in Appendix Table I. Standard errors, reported in parentheses, are calculated using the corresponding standard errors in Table III and Appendix Table I. In each panel, the first row reports the mean elasticity implied for the U.S. as a whole, while the second row reports the elasticity in the top bunching decile of ZIP-3-by-year cells. The identifying assumption in both cases is that the elasticity is zero in the bottom bunching decile. Column 1 reports the intensive margin elasticity required in a neoclassical model of frictionless optimization to generate the increase in EITC amounts around child birth estimated in column 1 of Table III for Panel A and column 1 of Appendix Table I for Panel B. Column 2 reports the elasticity in the phase-in range required to generate the increase in the phase-in EITC amounts estimated in column 5 of Table III for Panel A and column 4 of Appendix Table I for Panel B. Column 3 reports the elasticity in the phase-out range required to generate the increase in the phase-out EITC amounts estimated in column 6 of Table III for Panel A and column 5 of Appendix Table I for Panel B. Column 4 reports estimates of participation elasticities using the estimates reported in column 7 of Table III and column 6 of Appendix Table I. The top decile elasticities are calculated to match the increase in EITC amounts around child birth in decile 10 relative to decile 1. See the text for additional details on the calculation of these elasticities.

TABLE V

Impact of EITC on Wage Earnings Distribution of EITC-Eligible Households

|	Percent of EITC-Eligible Households Below Threshold			
	(1)	(2)	(3)	(4)
	50% of Poverty Line	100% of Poverty Line	150% of Poverty Line	200% of Poverty Line
No EITC Counterfactual	13.71%	31.91%	54.31%	77.27%
EITC with No Behavioral Response	9.40%	21.95%	42.14%	71.11%
EITC with Avg. Behavioral Response in U.S.	8.16%	21.00%	41.97%	71.29%
EITC with Top Decile Behavioral Response	6.15%	19.56%	41.99%	71.73%

Notes: This table presents CDF's of wage earnings distributions under various scenarios. Each column reports the CDF of the income distribution of EITC-eligible wage earners with dependents at various thresholds relative to the Federal Poverty Line (FPL). We calculate the FPL for each observation in our sample based on year, marital status and number of children. The first row shows statistics for the counterfactual wage earnings distribution if there were no EITC. To construct this distribution, we first estimate the causal impact of the EITC on wage earnings using the difference-in-differences estimator around child birth described in equation (7). We then subtract this estimate of the causal impact of the EITC from the CDF of the observed unconditional wage earnings distribution in our sample between 2000-2005. The second row recomputes the CDF in the first row after mechanically adding the EITC payments each household would receive based on its characteristics. The third row reports the observed CDF in our sample using the unconditional post-EITC wage earnings distribution. This row incorporates the effects of both mechanical transfers and behavioral responses to the EITC. The fourth row reports the counterfactual net earnings distribution if the level of information increased in all areas to that of neighborhoods in the highest decile of self-employed sharp bunching in our sample. We estimate this effect by recalculating the difference-in-differences estimate of the causal impact of the EITC using the top bunching decile instead of the full sample. We then add this causal effect back to the counterfactual distribution calculated in the first row and recompute EITC refund amounts.

APPENDIX TABLE I
Impacts of EITC on Total Earnings: Regression Estimates from Child Birth Design

| Dependent Variable:	Baseline Specification	With ZIP-3 Effects	Placebo Test: 3rd Child	Simulated Phase-in Credit	Simulated Phase-out Credit	Positive Total Earnings	Mean EITC x (Earn > 0)
	Simulated EITC Refund			Phase-in vs. Phase-out		Extensive Margin	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ZIP-3 Self-Emp. Sharp Bunching (%)	$44.2 (2.60)	$47.5 (0.99)	$2.1 (0.86)	$36.9 (2.39)	$7.3 (0.81)	0.97% (0.07)	$10.8 (0.72)
ZIP-3 by Post-Birth Fixed Effects		x					
Observations (millions)	29.96	29.96	10.07	29.96	29.96	29.96	29.96

Notes: This table replicates selected columns from Table III using total earnings (self-employment income plus wage earnings) to calculate the simulated EITC refund. See Table III for details on the variables and specifications.

FIGURE 1
Aggregate Earnings Distributions for EITC-Eligible Tax Filers

a) All Households with Children in 2008

b) Wage-Earners with Children in 2008
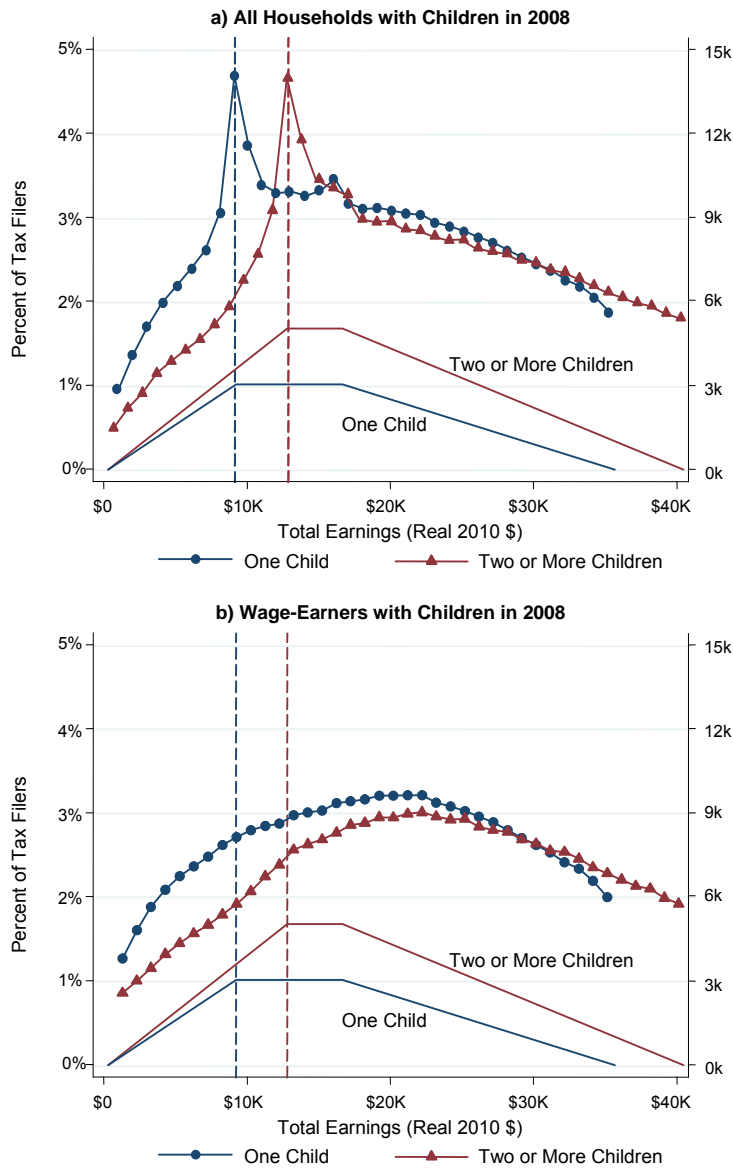
Notes: Panel A plots the distribution of total earnings for all individuals in our cross-sectional analysis sample in 2008, which includes primary tax filers who report one or more children and have income in the EITC-eligible range. This and all subsequent distributions are histograms with $1,000 bins centered around the first kink of the EITC schedule. Total earnings is the total amount of earnings used to calculate the EITC and is essentially the sum of wage earnings and self-employment income reported on form 1040. We plot separate distributions for households claiming one child and households claiming two or more children. Panel B repeats Panel A for wage earners, i.e. households who report no self-employment (Schedule C) income in 2008. Each panel also shows the EITC credit schedule for single filers with one and two or more children in 2008 (right scale). The dashed lines depict the income level that maximizes refunds net of other tax liabilities. Married households filing jointly face schedules with the same first kink point, but a plateau region extended by $3,000. In this and all subsequent figures, dollar values are scaled in 2010 real dollars using the IRS inflation adjustment.

FIGURE 2
Self-Employed Sharp Bunching Rates Across Neighborhoods

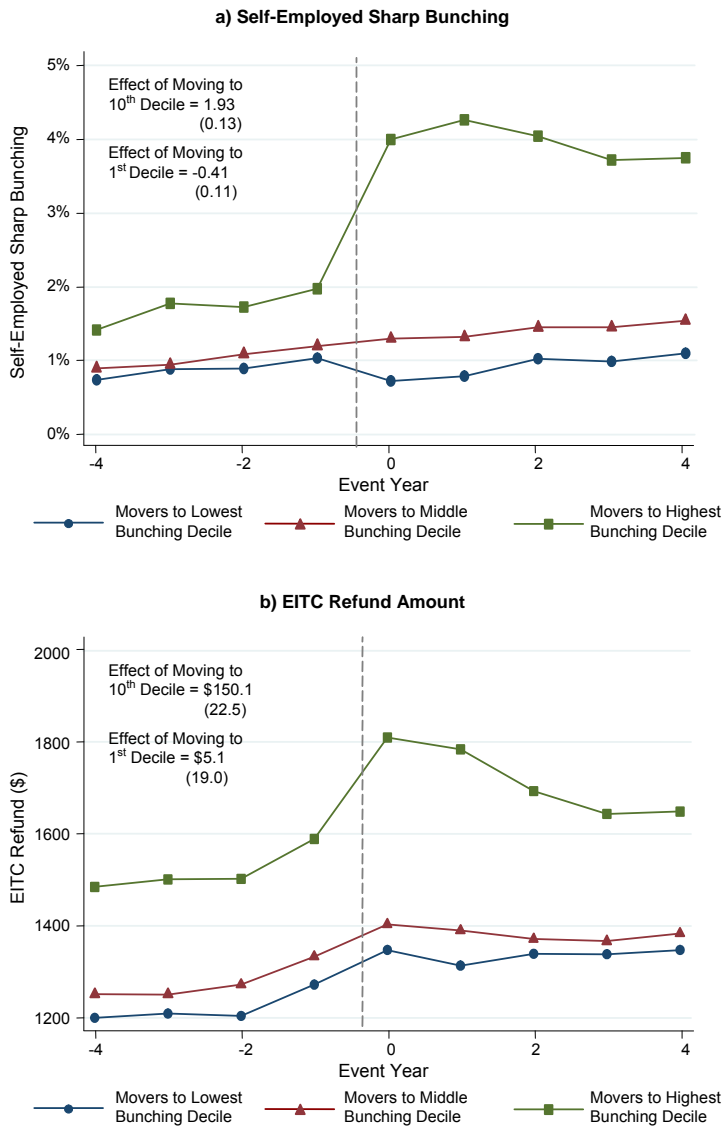|	
■	4.4 − 30.6%
■	3.2 − 4.4%
■	2.5 − 3.2%
■	2.2 − 2.5%
■	1.9 − 2.2%
■	1.6 − 1.9%
■	1.4 − 1.6%
■	1.2 − 1.4%
■	0.9 − 1.2%
■	0 − 0.9%
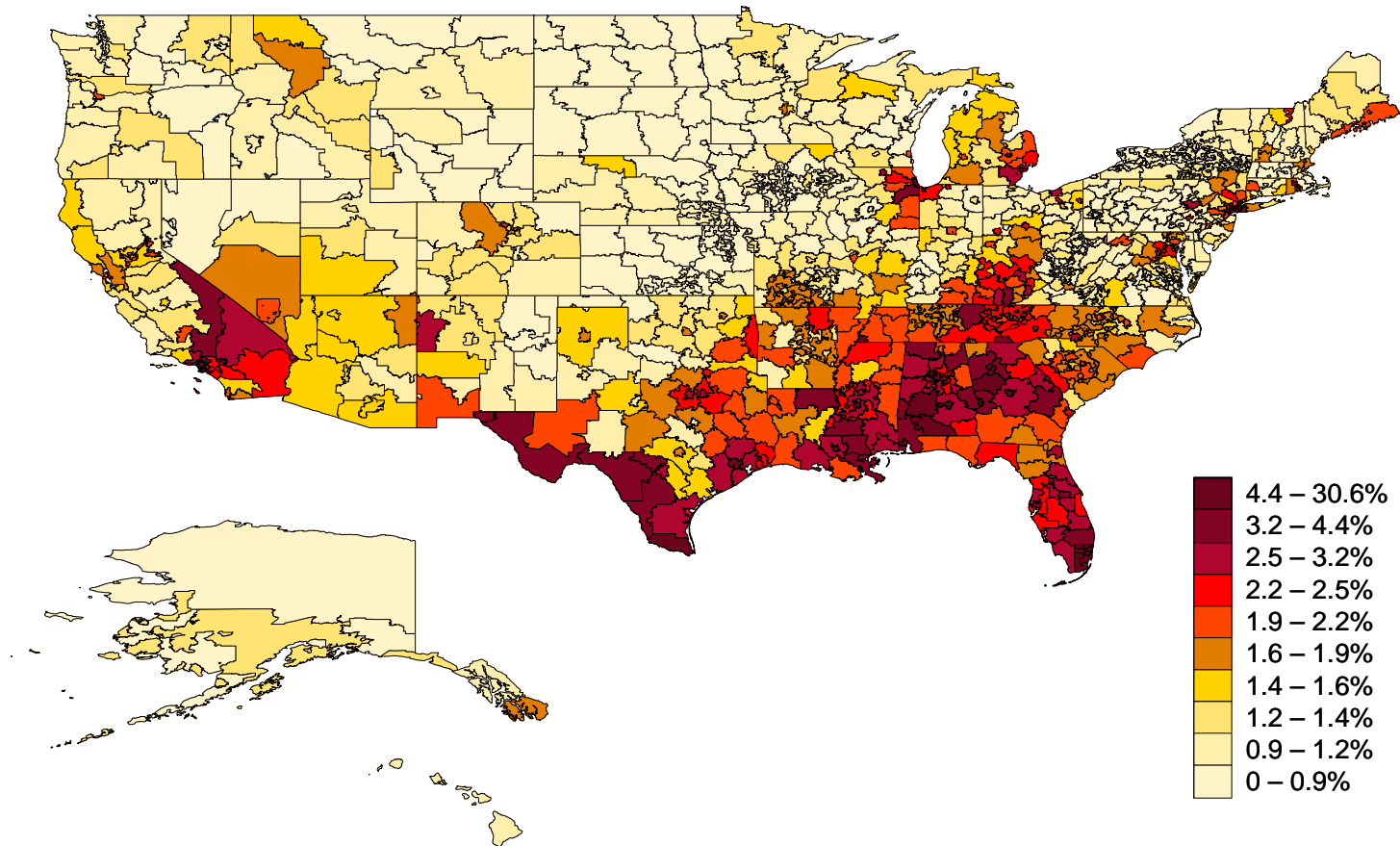
Notes: This figure plots sharp bunching rates by ZIP-3 in 2008. Self-employed sharp bunching is defined as the fraction of all EITC-eligible households with children in the cross-sectional sample whose total income falls within $500 of the first kink point and who have non-zero self-employment income. We divide the observations into deciles within the 2008 cross-sectional sample. Each decile is assigned a different color on the map, with darker shades representing higher levels of sharp bunching.

FIGURE 3
Earnings Distributions in Lowest vs. Highest Sharp Bunching Deciles
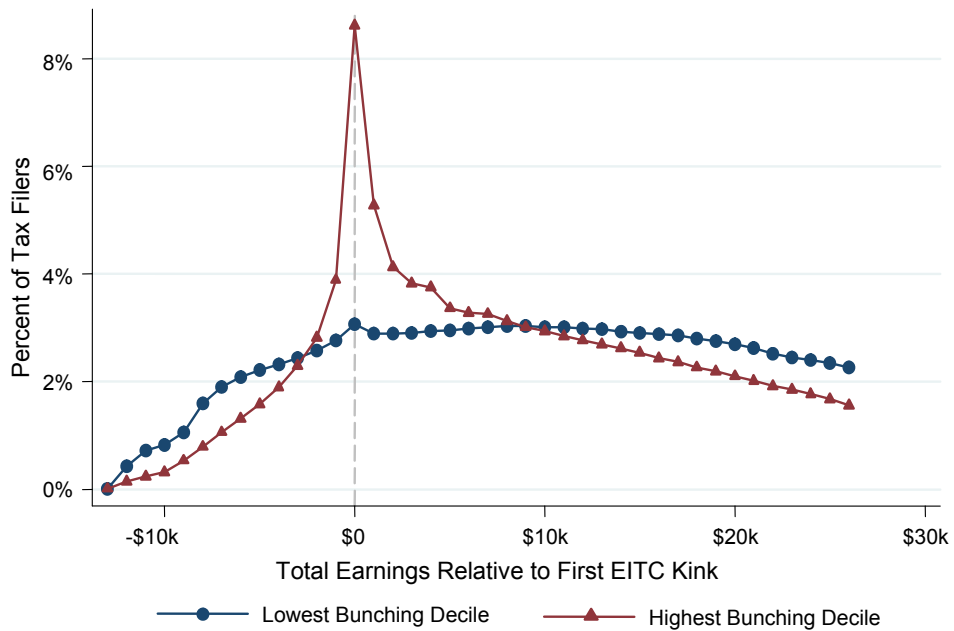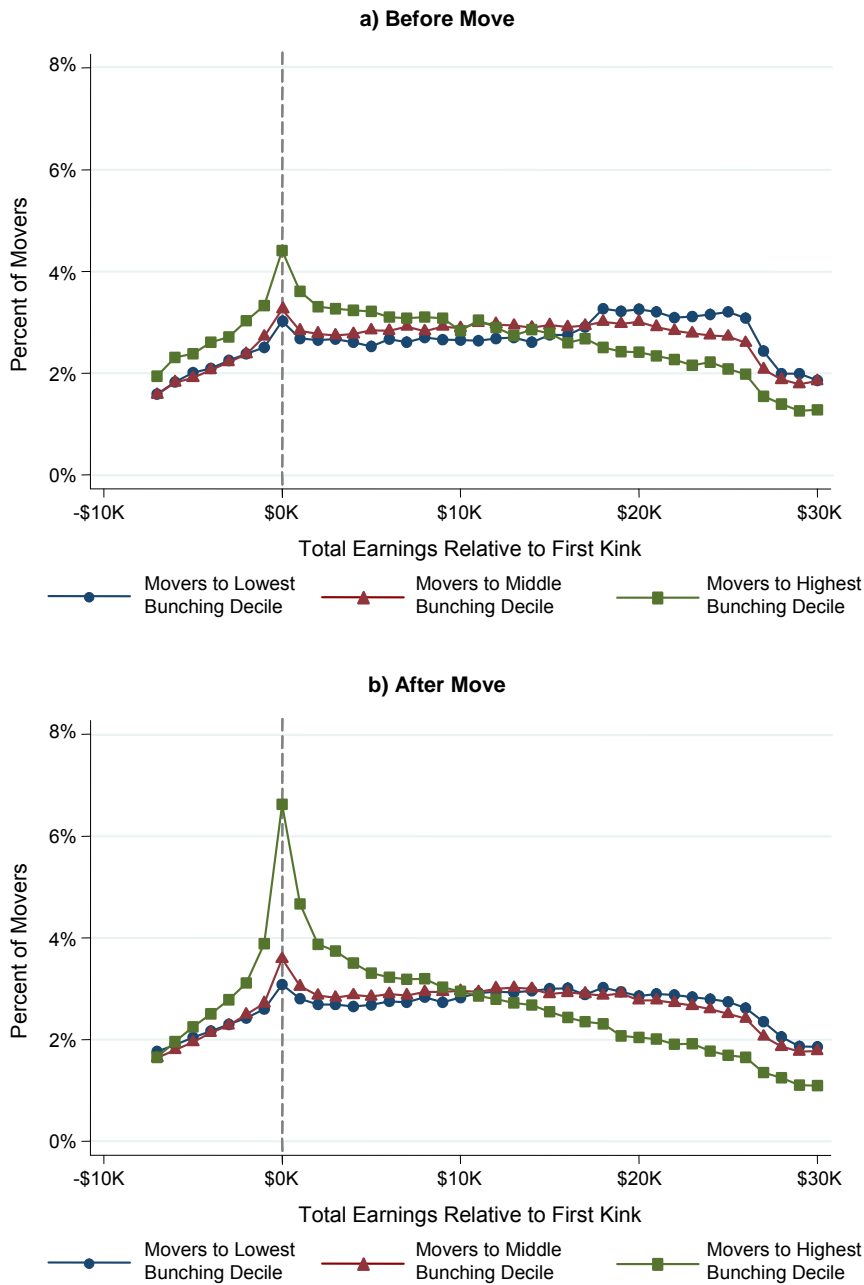

Notes: This figure plots the distribution of total earnings for individuals living in ZIP-3-by-year cells in the highest and lowest deciles of self-employed sharp bunching. Self-employed sharp bunching is defined as the percentage of EITC claimants with children in the ZIP-3-by-year cell who report total earnings within $500 of the first EITC kink and have non-zero self-employment income. We use all years in the cross-sectional analysis sample (1996-2009) in this figure. We divide the observations into deciles after pooling all years of the sample, so that the decile cut points remain fixed across years. The figure includes individuals with both 1 and 2+ children by plotting total earnings minus the first kink point of the relevant EITC schedule, so that 0 denotes the refund-maximizing point.

FIGURE 4
Event Studies of Movers

a) Self-Employed Sharp Bunching

Effect of Moving to
10th Decile = 1.93
(0.13)

Effect of Moving to
1st Decile = -0.41
(0.11)

Movers to Lowest Bunching Decile • Movers to Middle Bunching Decile ▲ Movers to Highest Bunching Decile ■

b) EITC Refund Amount

Effect of Moving to
10th Decile = $150.1
(22.5)

Effect of Moving to
1st Decile = $5.1
(19.0)

Movers to Lowest Bunching Decile • Movers to Middle Bunching Decile ▲ Movers to Highest Bunching Decile ■

Notes: Each panel plots an event study of individuals who move across ZIP-3s. We define event time as the calender year minus the year of the move, so year 0 is the year in which the individual moves. The figure is drawn using the movers sample, which includes all individuals in our core sample who move across ZIP-3s in any year between 2000 and 2005. If an individual moves more than once, we use only the first move. To construct the figure, we first define the degree of bunching for prior residents of ZIP-3 c in year t as the sharp bunching rate for individuals in the cross-sectional analysis sample living in ZIP-3 c in year $t-1$. We then divide the ZIP-3-by-year cells into ten deciles of prior residents' bunching rates by splitting the individual-level observations in the movers sample into ten equal-sized groups. Each figure plots outcomes for individuals who move from ZIP-3-by-year cells in the 5th decile to cells in the 1st, 5th, and 10th deciles. The outcome in Panel A is the rate of self-employed sharp bunching among the movers themselves. The outcome in Panel B is the mean EITC refund for the movers. In both panels, we include only individual-year observations in which the mover has one or more children and has total earnings in the EITC-eligible range. The coefficients and standard errors are estimated using difference-in-differences regression specifications comparing changes from year -1 to 0 for movers to the 10th or 1st deciles with changes for those moving to the 5th decile. See text for details. Standard errors are clustered at the ZIP-3-by-year of move level.

FIGURE 5
Total Earnings Distributions Before and After Move
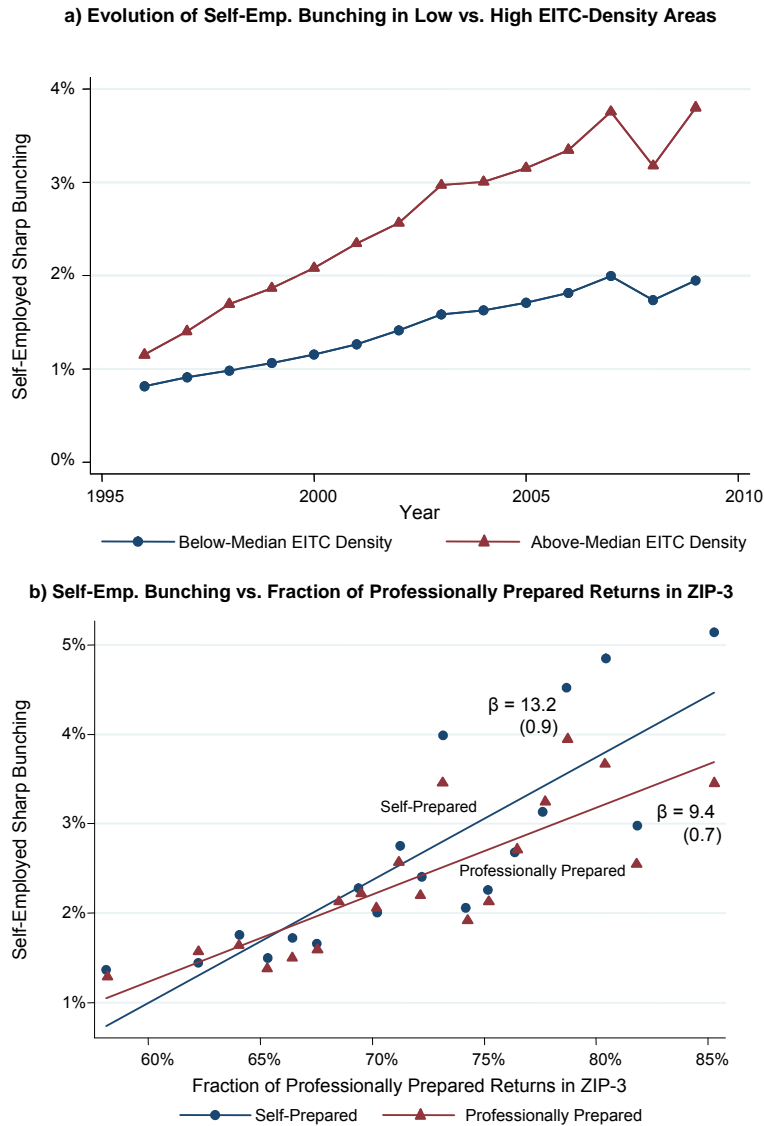
a) Before Move

b) After Move

Notes: These figures plot the distribution of total earnings before and after moving for the three groups of movers shown in Figure 4. Panel A shows the distribution of total earnings relative to the first kink point in the year before the move. Panel B repeats this exercise for the year of the move. As in Figure 2, we include individuals with both 1 and 2+ children by plotting total earnings minus the first kink point of the relevant EITC schedule, so that 0 denotes the refund-maximizing point. See the notes to Figure 4 for details on sample and decile definitions.

FIGURE 6
Impact of Moving to Neighborhoods with Lower vs. Higher Sharp Bunching
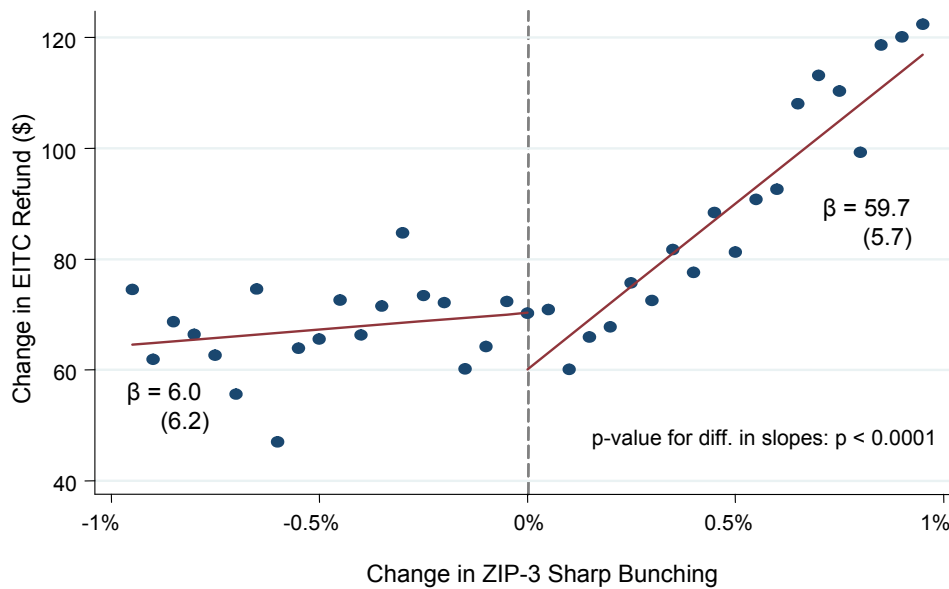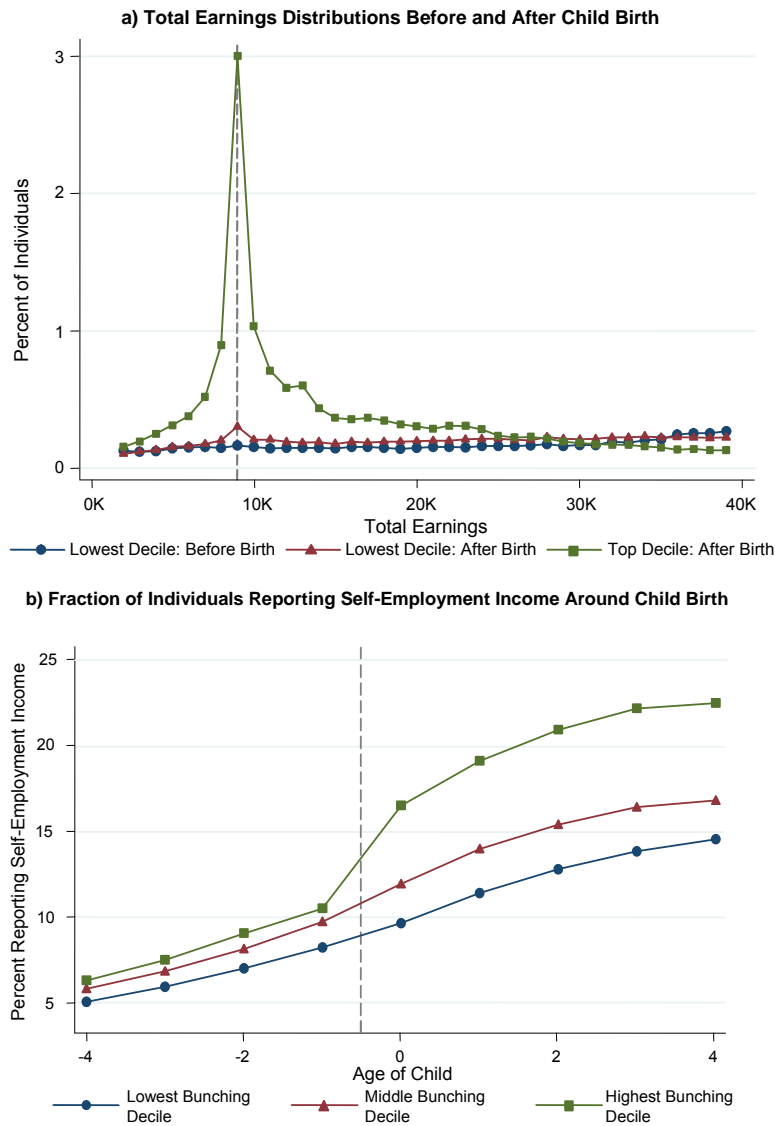

Notes: This figure plots changes in EITC refund amounts from the year before the move (event year -1 in Figure 4) to the year after the move (event year 0) vs. changes in the level of residents' sharp bunching across the old and new ZIP-3s. We define the change in ZIP-3 sharp bunching as the difference between bunching of prior residents of the ZIP-3 where the mover lives before the move and bunching of the ZIP-3 where the mover lives after the move. As in Figure 4, bunching for prior residents of ZIP-3 c in year t is defined as the sharp bunching rate in year t for individuals in the cross-sectional analysis sample living in ZIP-3 c in year $t-1$. Bunching after the move is defined as the sharp bunching rate in year t in the mover's new ZIP-3. To construct the figure, we group individuals into 0.05%-wide bins on changes in sharp bunching and then plot the means of the change in average EITC refund within each bin. The solid lines represent best-fit linear regressions estimated on the microdata separately for observations above and below 0. The estimated slopes are reported next to each line along with standard errors clustered by bin. See the notes to Figure 4 for further details on the sample definitions.

FIGURE 7
Correlates of Sharp Bunching

a) Evolution of Self-Emp. Bunching in Low vs. High EITC-Density Areas

b) Self-Emp. Bunching vs. Fraction of Professionally Prepared Returns in ZIP-3


Notes: Panel A plots sharp bunching rates by year for two groups: ZIP-3s with above-median and below-median EITC filer density. We calculate density as the number of EITC-eligible filers per square mile. We split ZIP-3s into two groups at the median based on their density in 1996 (weighting by the number of individuals in each ZIP-3), and then plot the average level of sharp bunching in each group over time. Panel B plots the relationship between bunching and the fraction of returns filed in each ZIP-3-by-year cell using third-party professional tax preparers. We define the use of a professional tax preparer as reporting either a Tax Preparer TIN (PTIN) or Tax Preparer EIN on Form 1040 and compute the fraction of returns using a professional tax preparer within each ZIP-3-by-year cell in our cross-sectional sample. To construct the plot in Panel B, we split the cross-sectional sample into twenty equal-sized bins based on the fraction of tax prepared returns. Within each bin, we then plot mean sharp bunching for two groups: filers who file their own returns and filers who themselves use a third-party preparer. Coefficients are from OLS regressions estimated at the ZIP-3-by-year level, weighted by the number of individuals in each cell, with standard errors reported in parentheses.

FIGURE 8
Impacts of Child Birth on Reported Self-Employment Income

a) Total Earnings Distributions Before and After Child Birth

Legend: Lowest Decile: Before Birth — Lowest Decile: After Birth — Top Decile: After Birth

b) Fraction of Individuals Reporting Self-Employment Income Around Child Birth

Legend: Lowest Bunching Decile — Middle Bunching Decile — Highest Bunching Decile

Notes: These figures are drawn using the child birth sample, which includes individuals from the core sample who give birth to their first child between 2000 and 2005. We classify individuals into deciles of sharp bunching based on the level of sharp bunching for residents of the ZIP-3 they inhabit in the year in which they have a child. Panel A includes only individuals with non-zero self employment income and plots the distribution of total earnings in the year before child birth for individuals in the lowest bunching decile, the distribution in the year of child birth for individuals in the lowest bunching decile, and the distribution in the year of child birth for individuals in the highest bunching decile. To simplify the figure, we omit a plot of pre-birth earnings for individuals in the highest bunching decile, since the distribution is similar to that of the lowest bunching decile, and in particular does not exhibit any sharp bunching around the first kink of the EITC schedule. Panel B plots an event study of the fraction of individuals in the child birth sample reporting non-zero self-employment income around child birth for individuals giving birth in 1st, 5th, and 10th decile ZIP-3s.
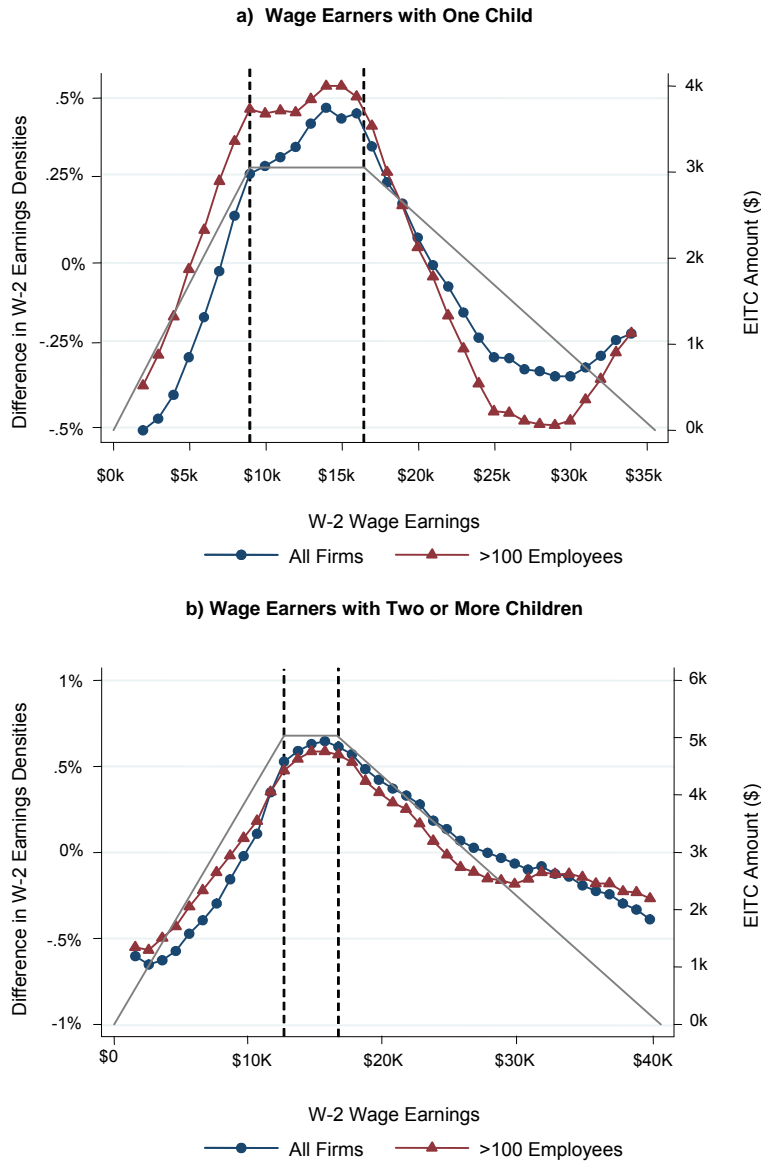
FIGURE 9
Wage Earnings Distributions in Lowest vs. Highest Bunching Deciles

a) Wage Earners with One Child

b) Wage Earners with Two or More Children
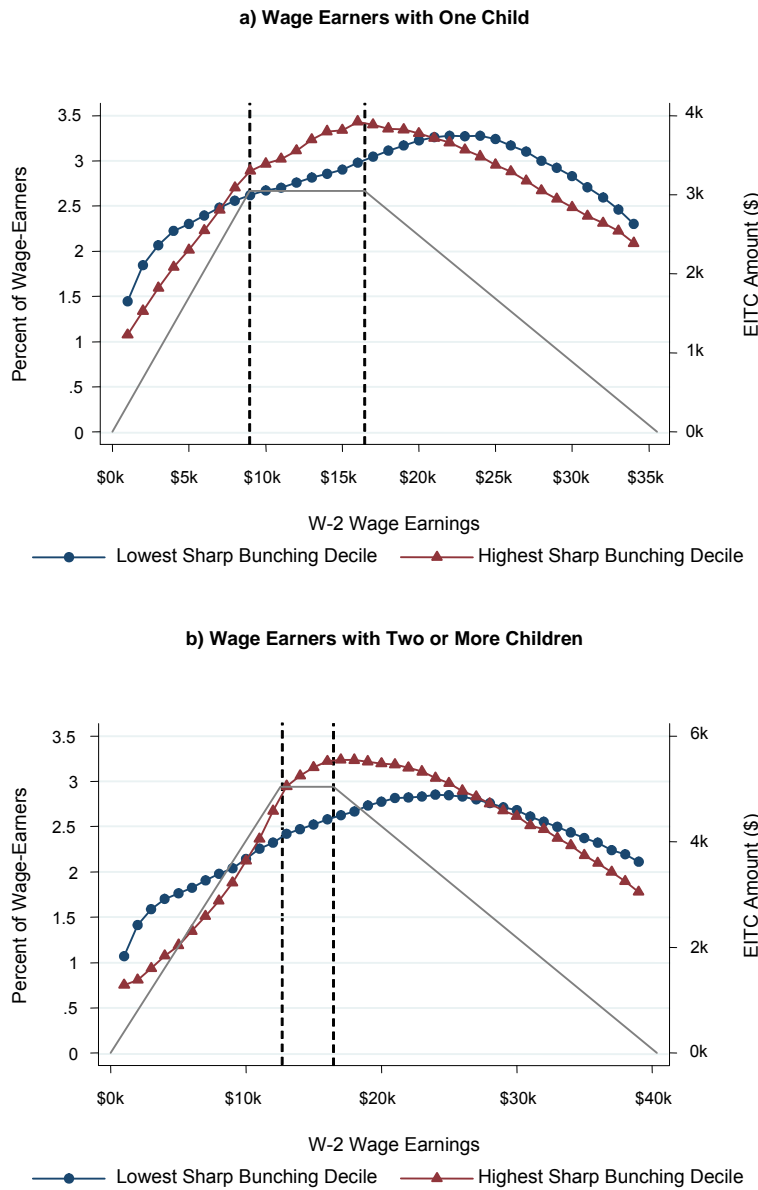
Notes: This figure plots W-2 wage earnings distributions for households without self-employment income using data from the cross-sectional sample from 1999-2009. The series in triangles includes individuals in ZIP-3-by-year cells in the highest self-employed sharp bunching decile, while the series in circles includes individuals in the lowest sharp bunching decile. Self-employed sharp bunching is defined as the percentage of EITC claimants with children in the ZIP-3-by-year cell who report total earnings within $500 of the first EITC kink and have non-zero self-employment income. We divide the observations in the pooled dataset covering 1999-2008 into deciles of sharp bunching, so that the decile cut points remain fixed across years. Panel A plots the distribution for households with one child; panel B plots the distribution for households with two or more children in 1999-2008 and exactly two children in 2009. The figures also show the relevant EITC schedule for single households in each panel (right scale); the schedule for married households has the same first kink point but has a plateau that is extended by an amount ranging from $1,000 in 2002 to $5,000 in 2009.
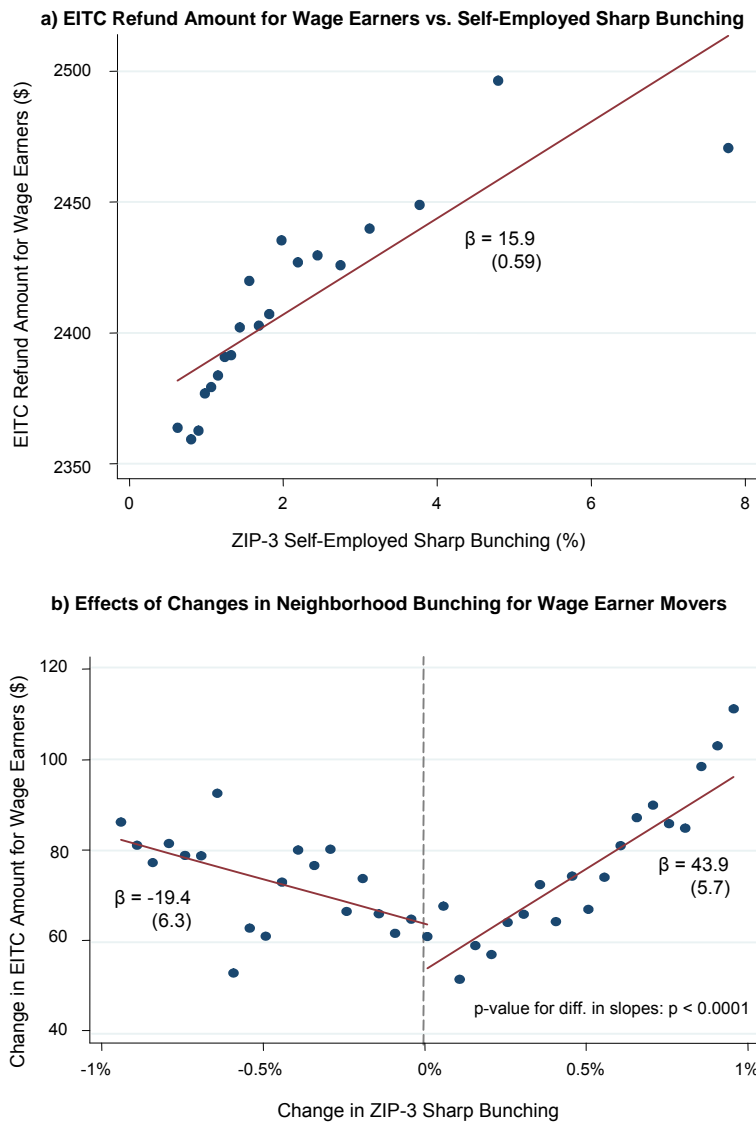
FIGURE 10

Differences in Wage Earnings Distributions: Lowest vs. Highest Bunching Deciles

a) Wage Earners with One Child

### b) Wage Earners with Two or More Children

Notes: This figure plots the difference in the W-2 wage-earnings distributions between the highest and lowest bunching deciles. The series in circles in Panel A is the difference between the two series plotted in Figure 9a; analogously, the series in circles in Panel B is the difference between the two series plotted in Figure 9b. The series in triangles replicate the analysis of the difference in earnings distributions, restricting attention to observations in the cross-sectional analysis sample in which all of the individual's W-2's came from firms that filed 100 or more W-2's in that year. The figures also show the relevant EITC schedule for single households in each panel (right scale); the schedule for married households has the same first kink point but has a plateau that is extended by an amount ranging from $1,000 in 2002 to $5,000 in 2009. See the notes to Figure 9 for further details.

FIGURE 11
Wage Earners' EITC Amounts vs. Self-Employed Sharp Bunching Rates

a) EITC Refund Amount for Wage Earners vs. Self-Employed Sharp Bunching

β = 15.9
(0.59)

b) Effects of Changes in Neighborhood Bunching for Wage Earner Movers

β = -19.4
(6.3)

β = 43.9
(5.7)

p-value for diff. in slopes: p < 0.0001

Notes: This figure plots the relationship between self-employed sharp bunching rates and EITC refund amounts for wage earners (those with no self-employment income). Panel A uses the cross-sectional analysis sample from 1999-2009; Panel B uses the movers sample. In both panels, we first calculate the EITC for each household. To construct Panel A, we split the observations into 20 equal-sized bins based on the rate of self-employed sharp bunching in the ZIP-3-by-year cell. We then plot the mean EITC refund vs. the mean sharp bunching rate in each bin. The best-fit line and coefficient are derived from an OLS regression of mean EITC refund amount in each ZIP-3-by-year cell on sharp bunching rates, weighted by the number of individuals in each cell. Panel B plots the relationship between change in EITC refund and change in neighborhood sharp bunching rate for movers who are wage earners. This figure replicates Figure 6, restricting the sample to wage earners and calculating the EITC refund based on W-2 wage earnings. See the notes to Figure 6 for more details on the construction of Panel B.

FIGURE 12
Wage Earnings Distributions Before and After Birth of First Child
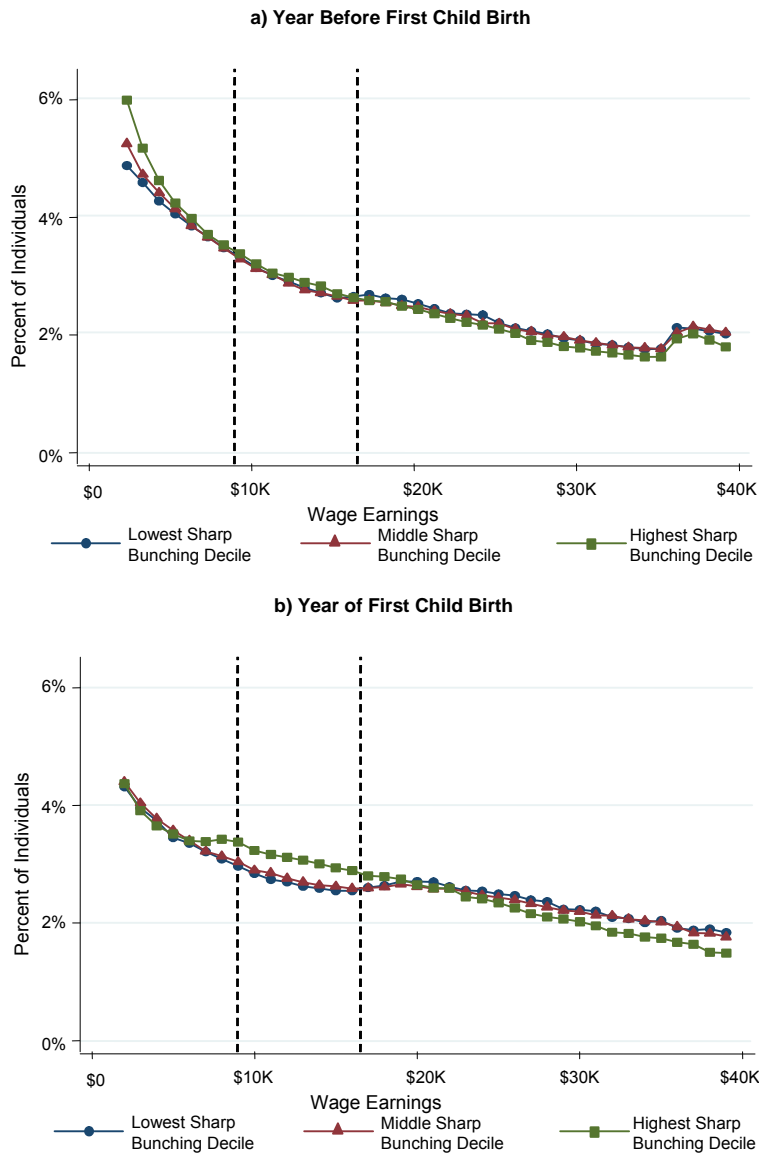
a) Year Before First Child Birth


b) Year of First Child Birth

Notes: These figures are drawn using the child birth sample, which includes individuals from the core sample who give birth to their first child between 2000 and 2005. We classify individuals into deciles of sharp bunching based on the level of sharp bunching for residents of the ZIP-3 they inhabit in the year in which they have a child. The figures only include wage-earners (those with no self-employment income) with positive W-2 earnings. Panel A plots W-2 wage earnings distributions in the year before child birth for individuals giving birth in ZIP-3-by-year cells in the 1st, 5th, and 10th deciles. Panel B replicates these distributions for the year of child birth. The dashed lines demarcate the beginning and end of the refund-maximizing plateau region of the EITC schedule for a single individuals with one child.

FIGURE 13
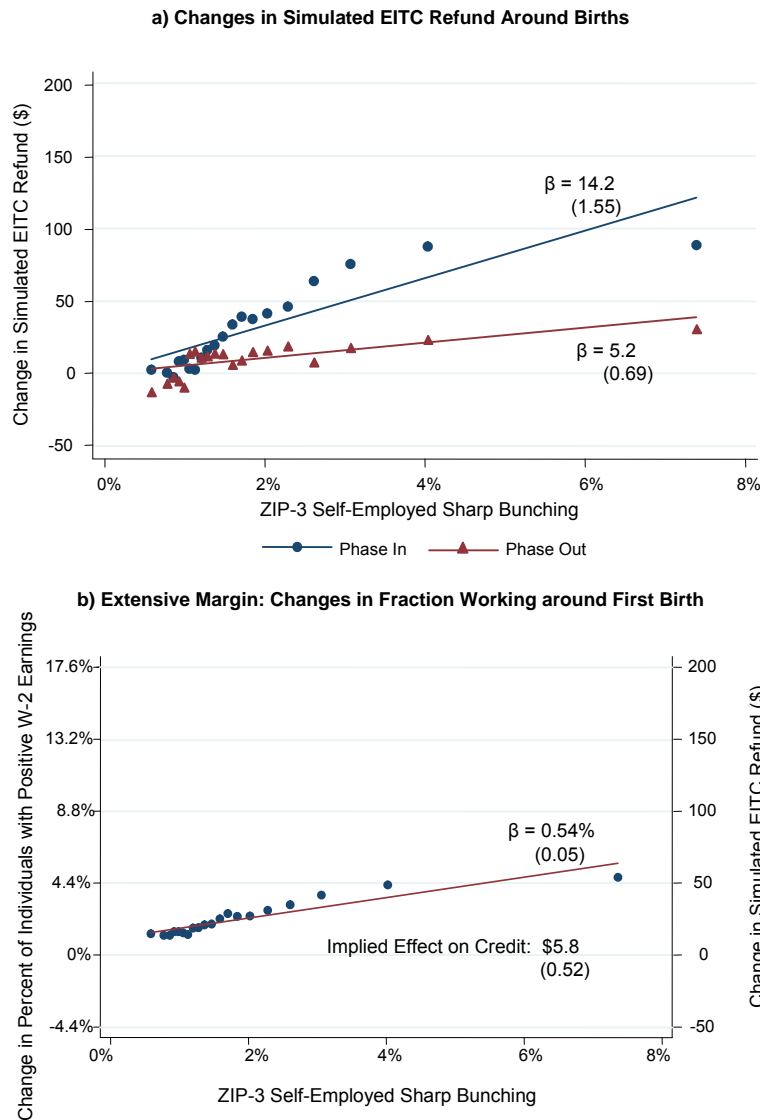## Event Study of Simulated EITC Around Birth of First Child
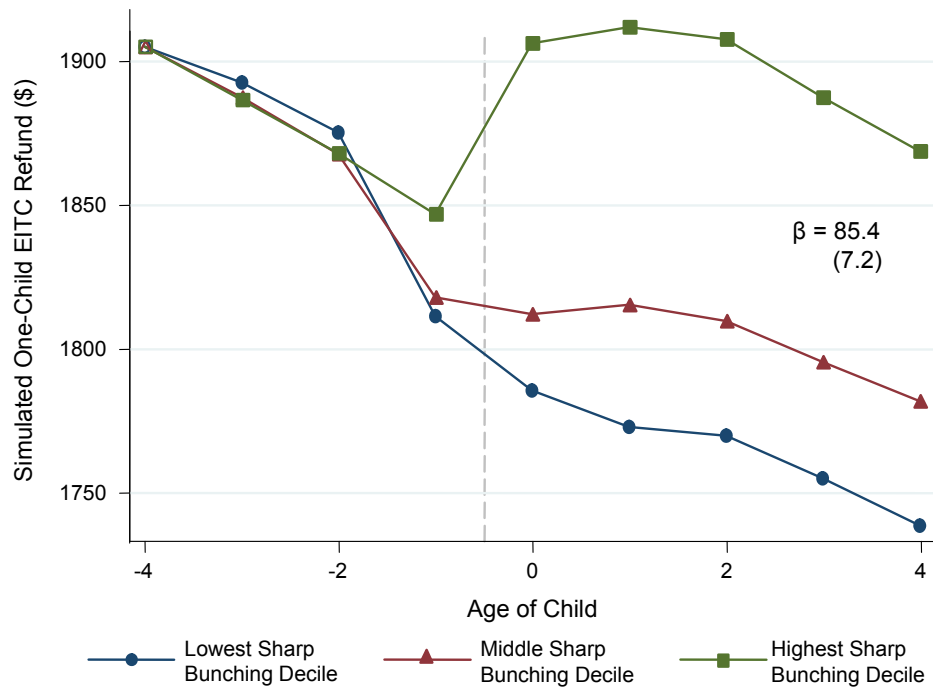
Notes: This figure plots an event study of the simulated EITC refund for wage earners around the year in which they have their first child. To calculate the simulated credit, we apply the one-child EITC schedule for single filers to total household W-2 earnings, regardless of the household's actual structure. The figure plots mean simulated credit amounts by event year for the exactly the same three groups as in Figure 12. For scaling purposes, we normalize the level of each series at the mean simulated credit in $t = -4$; that is, we subtract the decile-specific mean in $t = -4$ and add back the mean simulated EITC across the three deciles in $t = -4$ to all observations. The coefficient compares changes in the simulated credit amount from year -1 to 0 across the highest and lowest bunching deciles, estimated using a difference-in-differences regression specification as described in the text. The standard error, reported in parentheses, is clustered at the ZIP-3-by-birth-year level. See the notes to Figure 12 for sample and bunching decile definitions.

FIGURE 14
Changes in EITC Refund Amounts Around Child Birth vs. Sharp Bunching Rates

a) Wage Earners Only

**b) Full Sample, with EITC Amounts based on Wage Earnings**
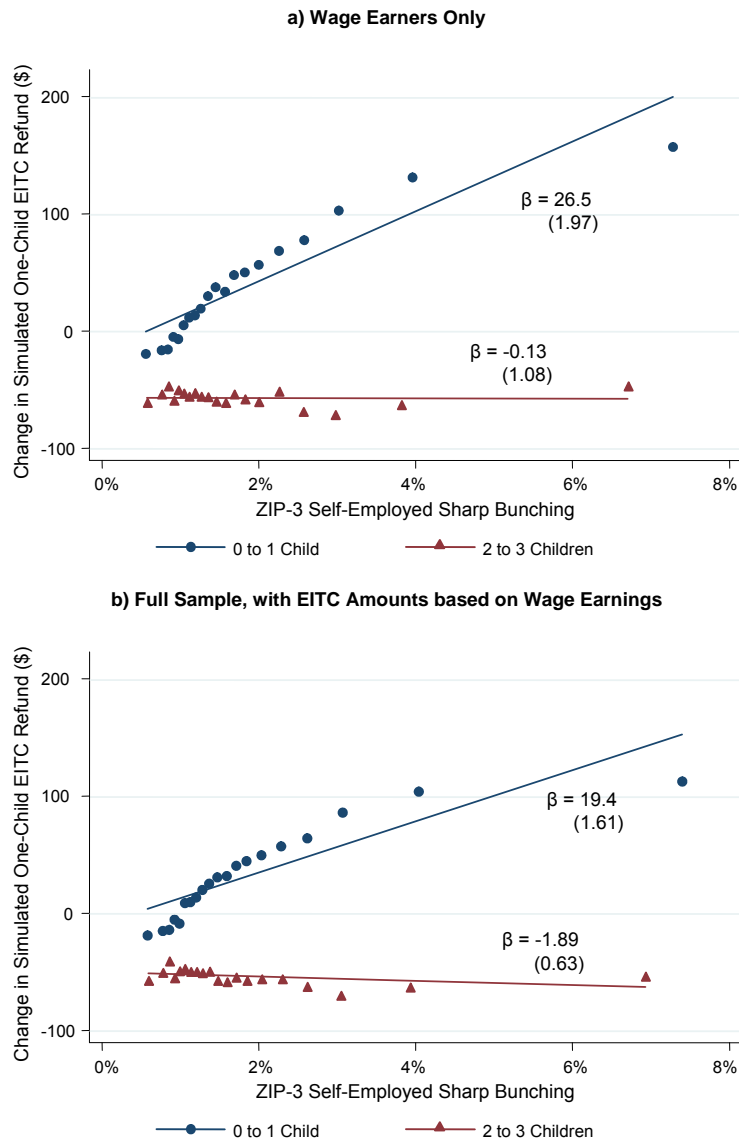
Notes: These figures plot changes in simulated EITC refund from the year before to the year of child birth (year -1 to year 0 in Figure 13) vs. the self-employed sharp bunching rate in the individual's ZIP-3 in the year of birth. Panel A includes only individuals in the child birth sample without self-employment income; Panel B includes all individuals in the child birth sample. In both panels we apply the one-child EITC schedule for single filers to total household W-2 earnings, regardless of the household's actual structure and self-employment income, to calculate the simulated credit. The series in circles plots changes in simulated one-child EITC around the birth of the first child; the series in triangles plots changes in simulated one-child EITC around the birth of the third child. To construct the "0 to 1 Child" series, we split the observations with first births into twenty equal-sized bins based on the degree of self-employed sharp bunching in the individual's ZIP-3-by-birth-year cell. Within each bin, we then calculate the mean change in simulated EITC from the year before to the year of the birth and plot this mean change against the sharp bunching rate. The "2 to 3 Child" series repeats this procedure for all third births (i.e, where the individual claimed two children the year before), once again using the one-child EITC schedule for single filers to calculate the simulated EITC credit. We estimate the best-fit lines and slopes using an OLS regression of the change in simulated credit on sharp bunching in the individual data, with standard errors clustered at the ZIP-3-by-birth-year level. See the notes to Figure 12 for further details on the child birth sample.

FIGURE 15
Phase-In, Phase-Out, and Extensive Margin Responses

a) Changes in Simulated EITC Refund Around Births

β = 14.2
(1.55)

β = 5.2
(0.69)

X-axis: ZIP-3 Self-Employed Sharp Bunching
Y-axis: Change in Simulated EITC Refund ($)

Legend: ● Phase In ▲ Phase Out

b) Extensive Margin: Changes in Fraction Working around First Birth

β = 0.54%
(0.05)

Implied Effect on Credit: $5.8
(0.52)

Left Y-axis: Change in Percent of Individuals with Positive W-2 Earnings
Right Y-axis: Change in Simulated EITC Refund ($)
X-axis: ZIP-3 Self-Employed Sharp Bunching

Notes: This figure decomposes the EITC response to the birth of a first child into the phase-in, phase-out and extensive margin responses. To do so, we replicate the "0 to 1 Child" series in Figure 14b, replacing the simulated EITC variable with other measures. Panel A distinguishes phase-in and phase-out responses. To calculate the phase-in response, we calculate the simulated credit using the schedule depicted in Appendix Figure 4a instead of the actual EITC schedule. For the phase-out response, we use the schedule depicted in Appendix Figure 4b instead. Panel B replaces the simulated EITC schedule with an indicator for positive W-2 wage earnings. We translate the extensive margin impact to an implied effect on EITC amounts by assuming that new workers earn the average EITC refund conditional on working in our sample ($1,075). The right scale in Panel B is chosen to match the scale of Panel A so that the size of the extensive margin response is scaled in the same units. The best-fit lines and standard errors are estimated as in Figure 14.
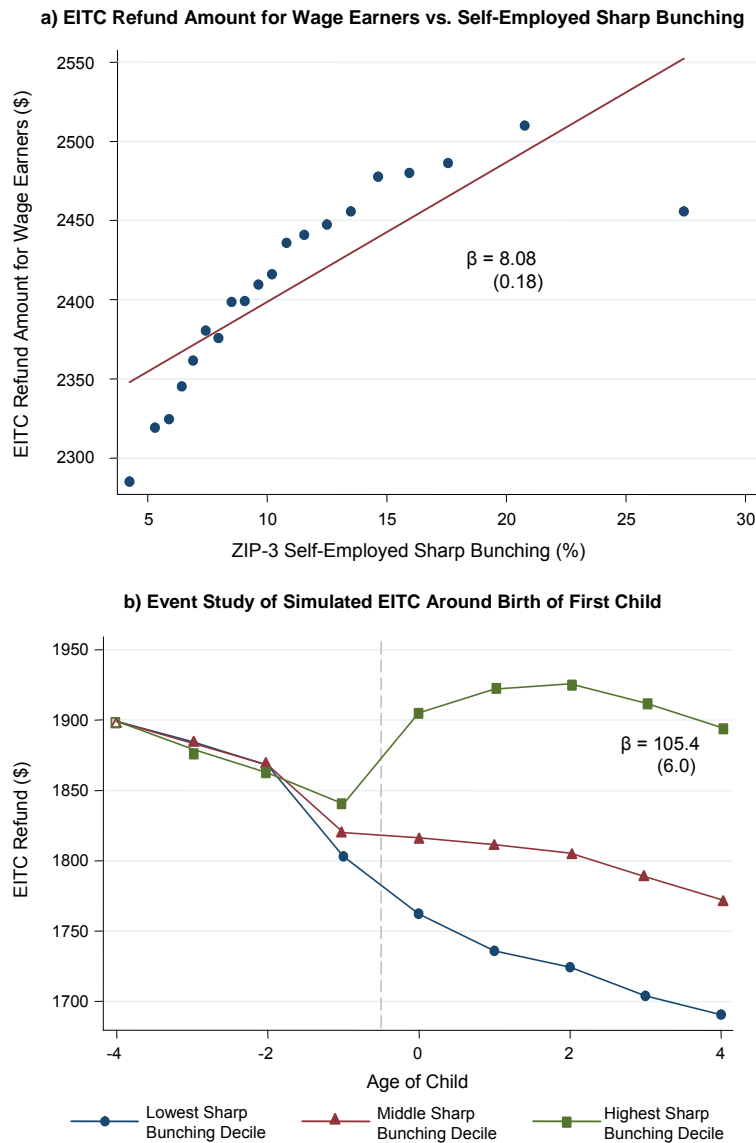
APPENDIX FIGURE 1
EITC Refund Schedule vs. Total Tax Liabilities for Single Filers with One Child
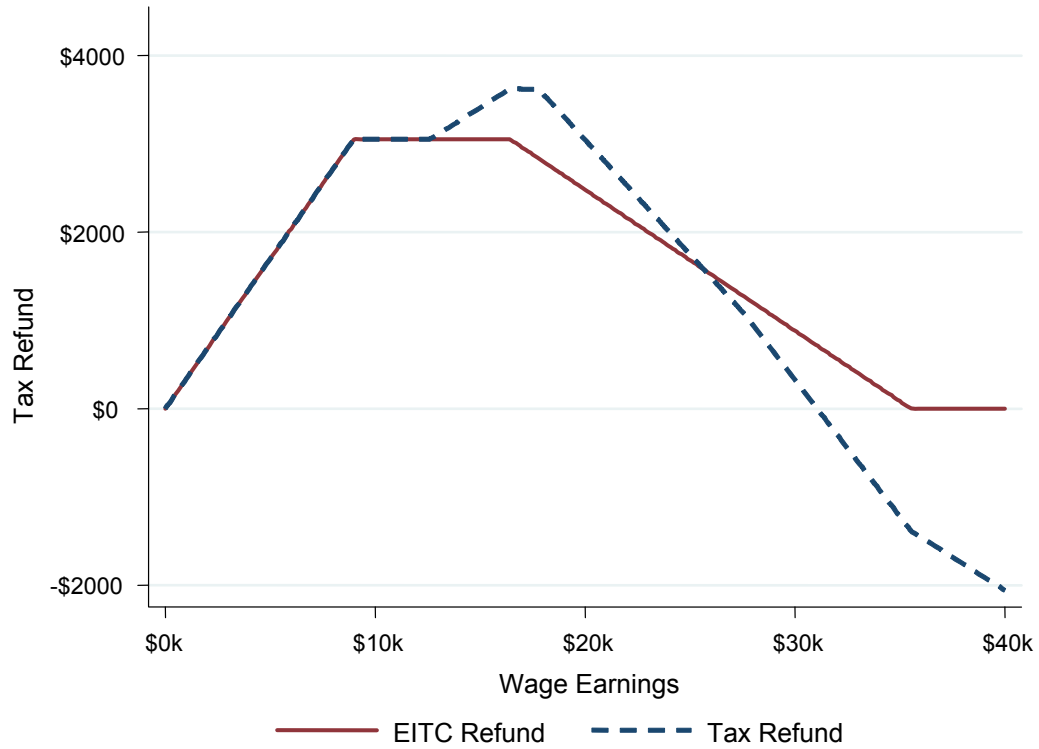
Notes: This figure plots the EITC refund and total tax refund for head-of-household filers with one dependent between 2002 and 2008. All monetary values are in 2010 dollars, indexed using the IRS inflation adjustment. The total tax refund includes the EITC and the Child Tax Credit (including the Additional Child Tax Credit) minus federal income taxes (but excluding payroll taxes). Negative values of the total tax refund indicate net tax liabilities.

APPENDIX FIGURE 2
Results with Alternative Measure of Sharp Bunching

a) EITC Refund Amount for Wage Earners vs. Self-Employed Sharp Bunching

β = 8.08
(0.18)

b) Event Study of Simulated EITC Around Birth of First Child

β = 105.4
(6.0)

Lowest Sharp Bunching Decile   Middle Sharp Bunching Decile   Highest Sharp Bunching Decile

Notes: This figure reproduces Figures 11a and 13 using an alternative definition of sharp bunching. Here, we define sharp bunching as the fraction of self-employed individuals in the ZIP-3-by-year cell who report income within $500 of the refund-maximizing kink. This definition differs from our baseline definition because we use the number of individuals with non-zero self-employment income in the denominator rather than the total number of individuals in the cross-sectional sample. In Panel A, we replace the baseline measure of sharp bunching with the alternative measure on the x-axis and reconstruct Figure 11a. To compare the coefficient in Panel A to that in Figure 11a, one must multiply the coefficient by 5.2 to account for the larger standard deviation of the alternative measure of sharp bunching. In Panel B, we define the sharp bunching deciles using the new measure and replicate Figure 13. The coefficient in Panel B can be compared directly with the coefficient in Figure 13.

APPENDIX FIGURE 3
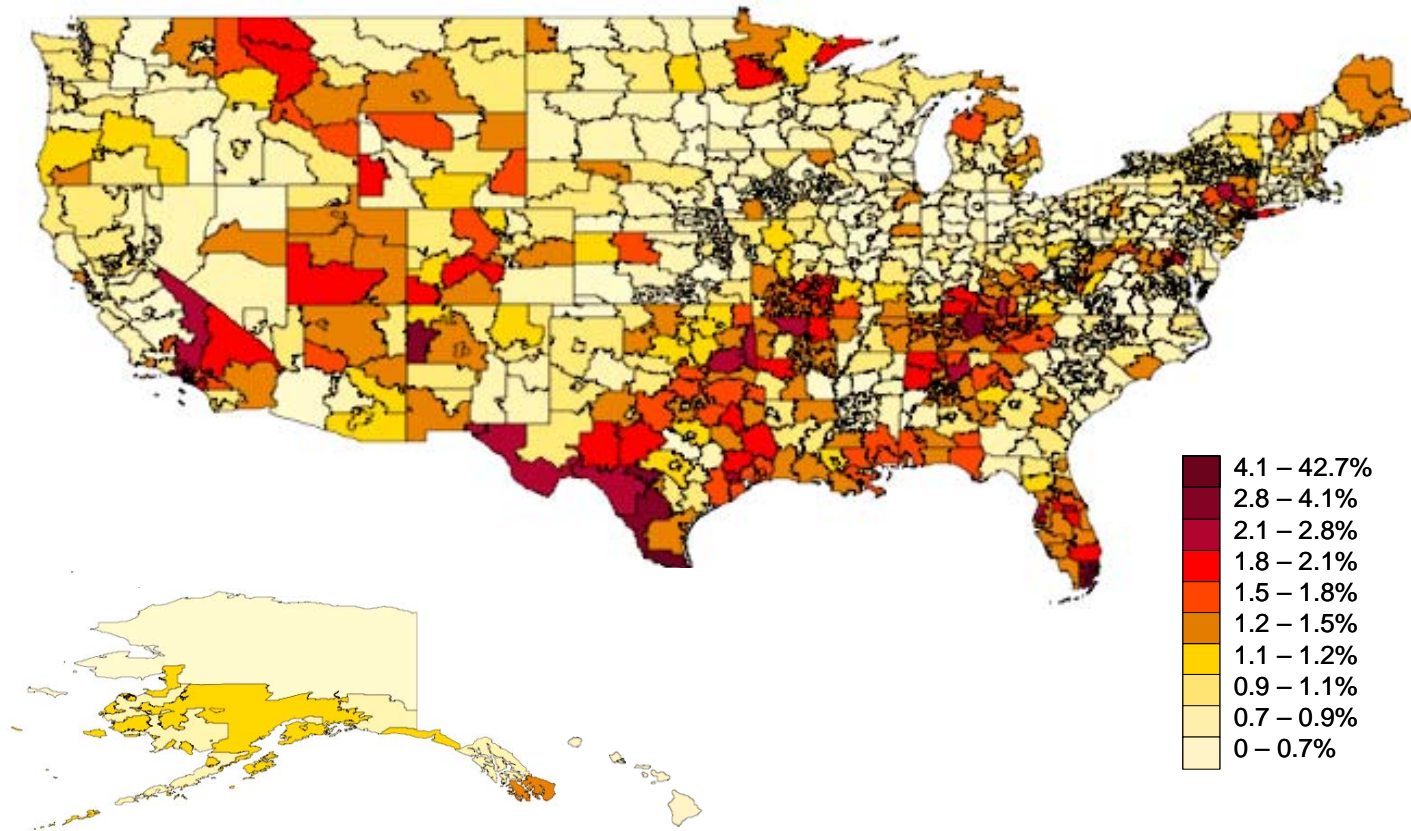## Self-Employed Sharp Bunching Rates Across Neighborhoods, 1996-2008

a) Self-Employed Sharp Bunching in 1996

■	4.1 – 42.7%
■	2.8 – 4.1%
■	2.1 – 2.8%
■	1.8 – 2.1%
■	1.5 – 1.8%
■	1.2 – 1.5%
■	1.1 – 1.2%
■	0.9 – 1.1%
■	0.7 – 0.9%
■	0 – 0.7%
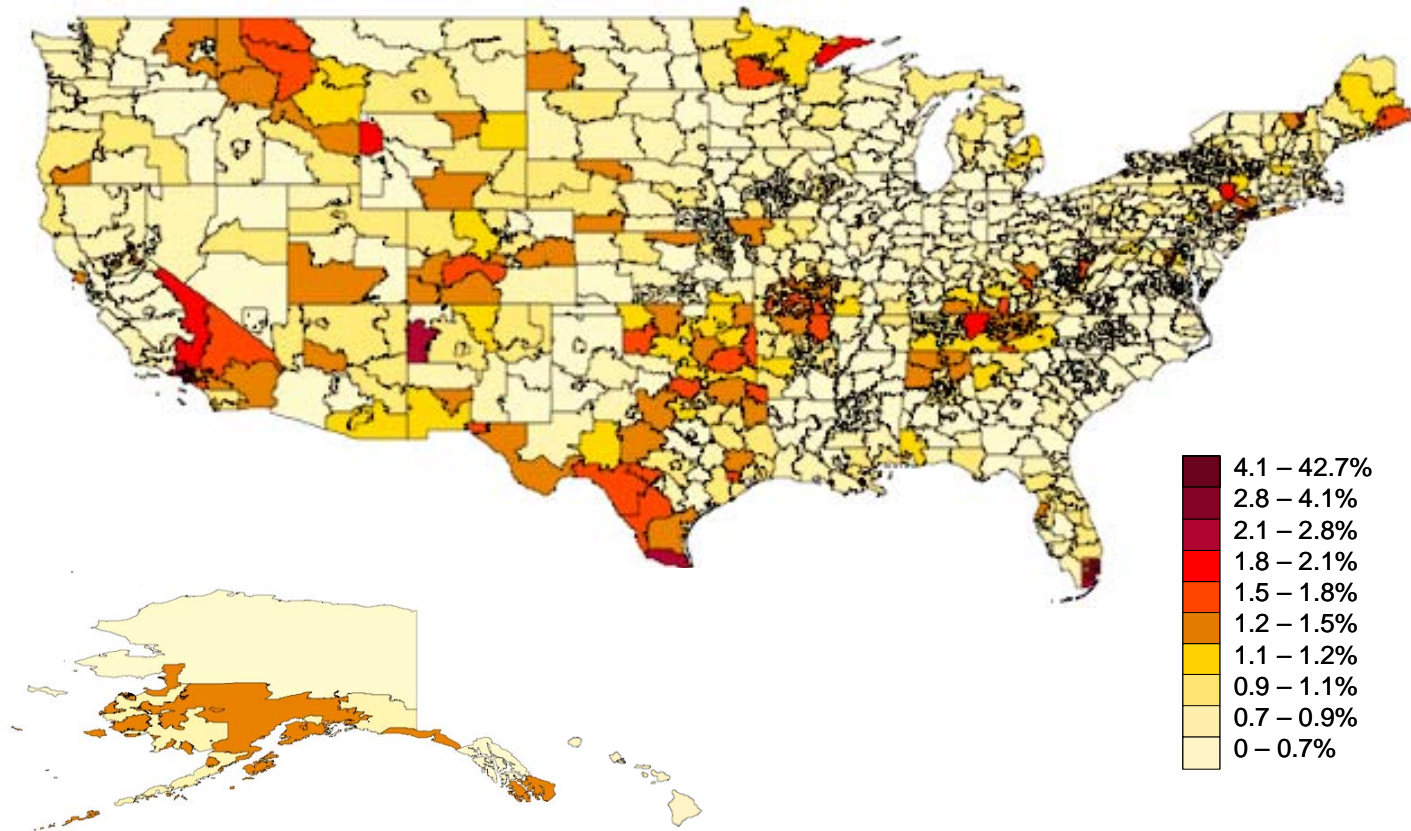
Notes: This figure plots sharp bunching rates by ZIP-3 in 1996. Self-employed sharp bunching is defined as the fraction of all EITC-eligible households with children in the cross-sectional sample whose total income falls within $500 of the first kink point and who have non-zero self-employment income. We divide the observations into deciles after pooling all years of the sample, so that the decile cut points remain fixed across years. Each decile is assigned a different color on the map, with darker shades representing higher levels of sharp bunching.

APPENDIX FIGURE 3
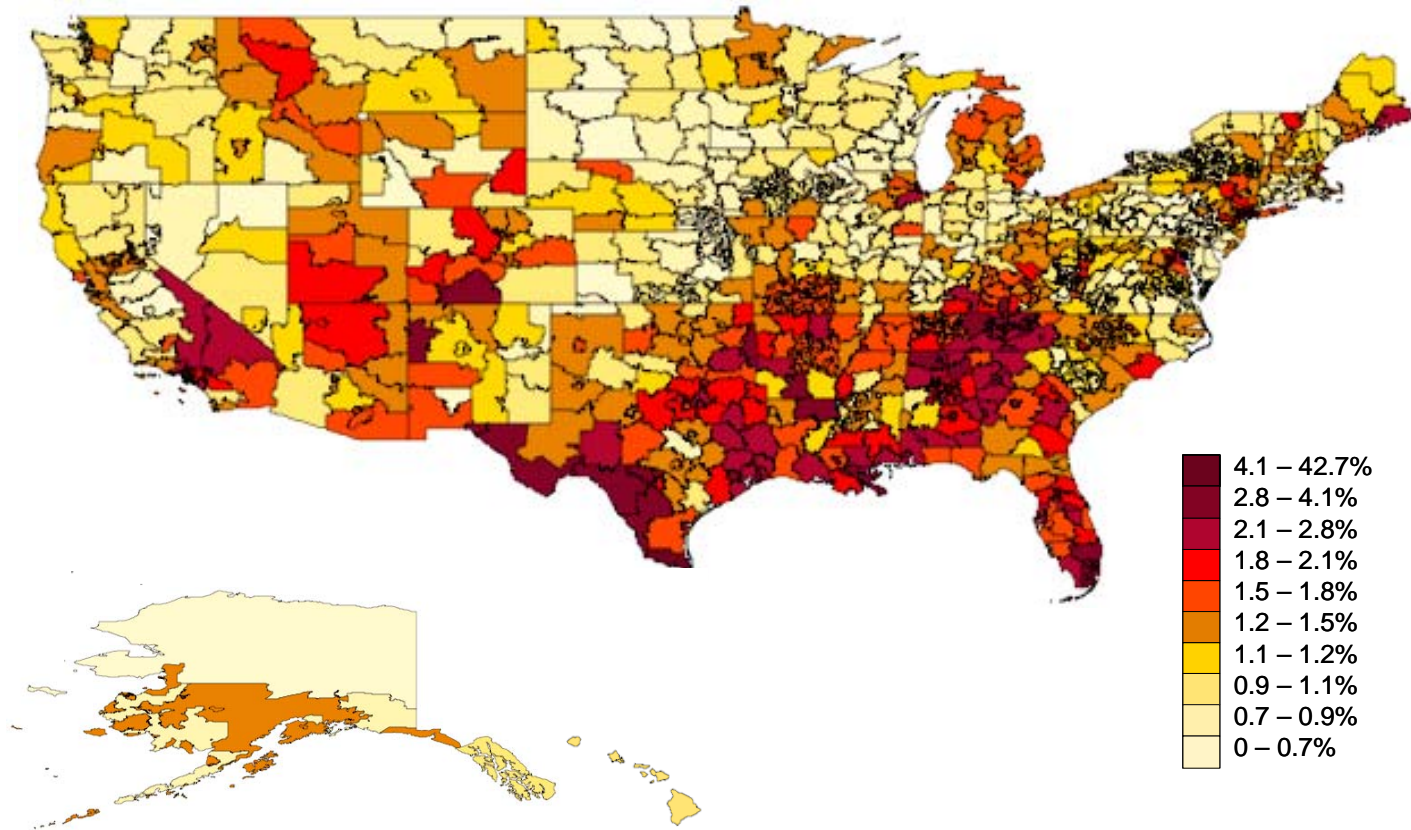## Self-Employed Sharp Bunching Rates Across Neighborhoods, 1996-2008
b) Self-Employed Sharp Bunching in 1999

■	4.1 – 42.7%
■	2.8 – 4.1%
■	2.1 – 2.8%
■	1.8 – 2.1%
■	1.5 – 1.8%
■	1.2 – 1.5%
■	1.1 – 1.2%
■	0.9 – 1.1%
■	0.7 – 0.9%
■	0 – 0.7%

Notes: This figure replicates Panel A for the year 1999. Self-employed sharp bunching is defined as the fraction of all EITC-eligible households with children in the cross-sectional sample whose total income falls within $500 of the first kink point and who have non-zero self-employment income. We divide the observations into deciles after pooling all years of the sample, so that the decile cut points remain fixed across years. Each decile is assigned a different color on the map, with darker shades representing higher levels of sharp bunching.

APPENDIX FIGURE 3
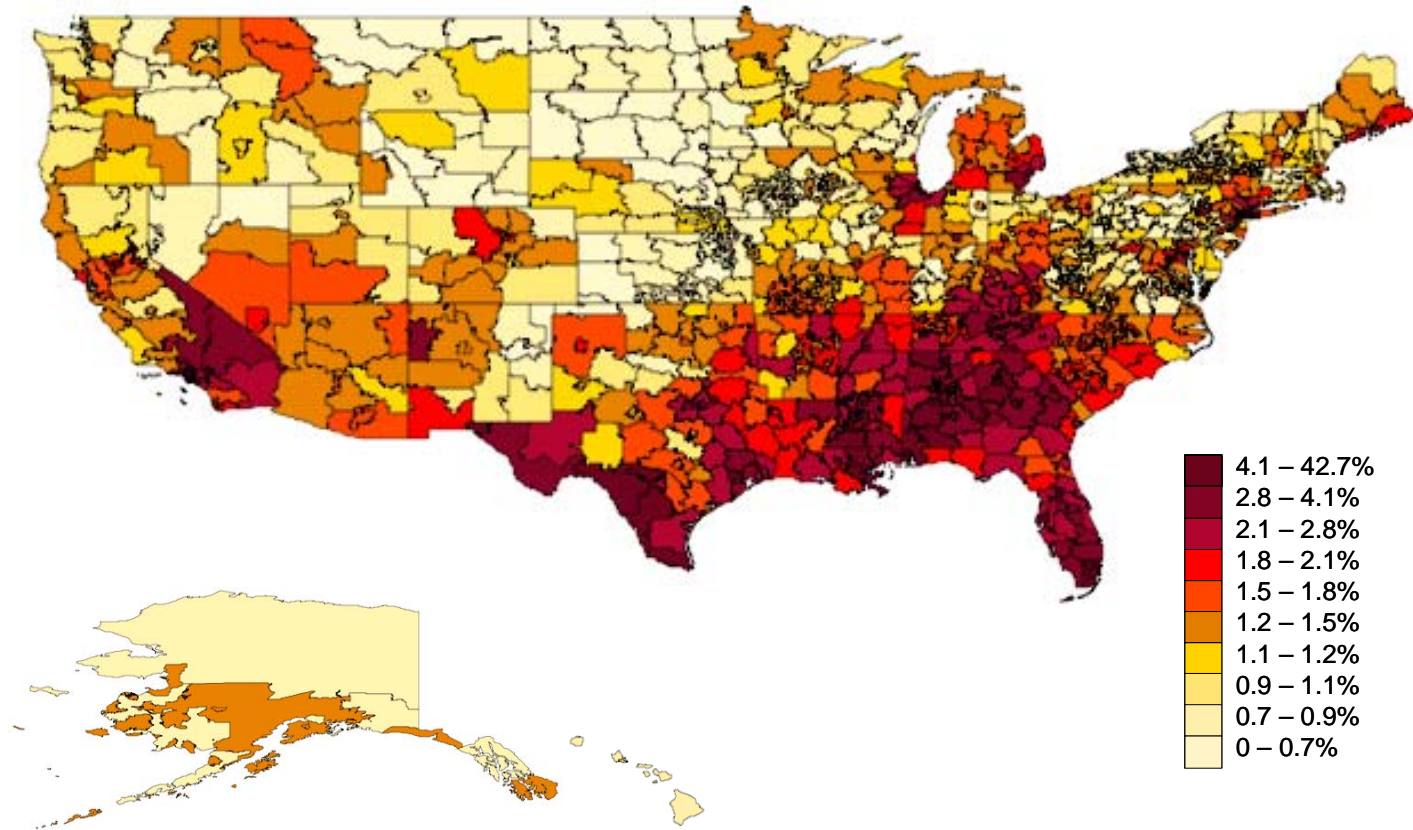## Self-Employed Sharp Bunching Rates Across Neighborhoods, 1996-2008
c) Self-Employed Sharp Bunching in 2002

Legend:
- 4.1 – 42.7%
- 2.8 – 4.1%
- 2.1 – 2.8%
- 1.8 – 2.1%
- 1.5 – 1.8%
- 1.2 – 1.5%
- 1.1 – 1.2%
- 0.9 – 1.1%
- 0.7 – 0.9%
- 0 – 0.7%

Notes: This figure replicates Panel A for the year 2002. Self-employed sharp bunching is defined as the fraction of all EITC-eligible households with children in the cross-sectional sample whose total income falls within $500 of the first kink point and who have non-zero self-employment income. We divide the observations into deciles after pooling all years of the sample, so that the decile cut points remain fixed across years. Each decile is assigned a different color on the map, with darker shades representing higher levels of sharp bunching.

APPENDIX FIGURE 3
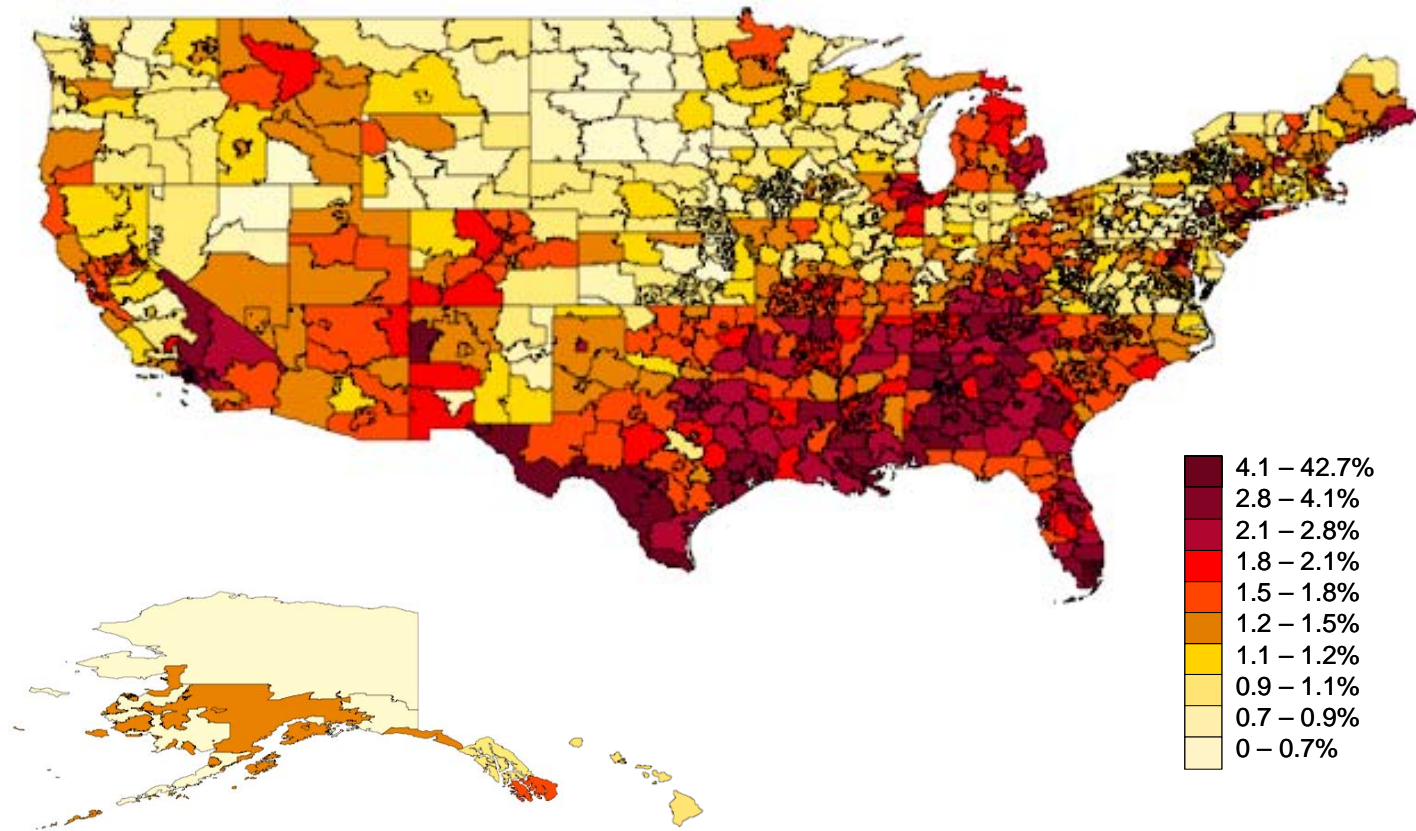## Self-Employed Sharp Bunching Rates Across Neighborhoods, 1996-2008
d) Self-Employed Sharp Bunching in 2005

■	4.1 – 42.7%
■	2.8 – 4.1%
■	2.1 – 2.8%
■	1.8 – 2.1%
■	1.5 – 1.8%
■	1.2 – 1.5%
■	1.1 – 1.2%
■	0.9 – 1.1%
■	0.7 – 0.9%
■	0 – 0.7%

Notes: This figure replicates Panel A for the year 2005. Self-employed sharp bunching is defined as the fraction of all EITC-eligible households with children in the cross-sectional sample whose total income falls within $500 of the first kink point and who have non-zero self-employment income. We divide the observations into deciles after pooling all years of the sample, so that the decile cut points remain fixed across years. Each decile is assigned a different color on the map, with darker shades representing higher levels of sharp bunching.

APPENDIX FIGURE 3
Self-Employed Sharp Bunching Rates Across Neighborhoods, 1996-2008
e) Self-Employed Sharp Bunching in 2008

■	4.1 − 42.7%
■	2.8 − 4.1%
■	2.1 − 2.8%
■	1.8 − 2.1%
■	1.5 − 1.8%
■	1.2 − 1.5%
■	1.1 − 1.2%
■	0.9 − 1.1%
■	0.7 − 0.9%
■	0 − 0.7%

Notes: This figure replicates Panel A for the year 2008. Self-employed sharp bunching is defined as the fraction of all EITC-eligible households with children in the cross-sectional sample whose total income falls within $500 of the first kink point and who have non-zero self-employment income. We divide the observations into deciles after pooling all years of the sample, so that the decile cut points remain fixed across years. Each decile is assigned a different color on the map, with darker shades representing higher levels of sharp bunching.

APPENDIX FIGURE 4
Simulated Phase-In and Phase-Out Credit Schedules

a) Simulated Phase-In Credit

b) Simulated Phase-Out Credit
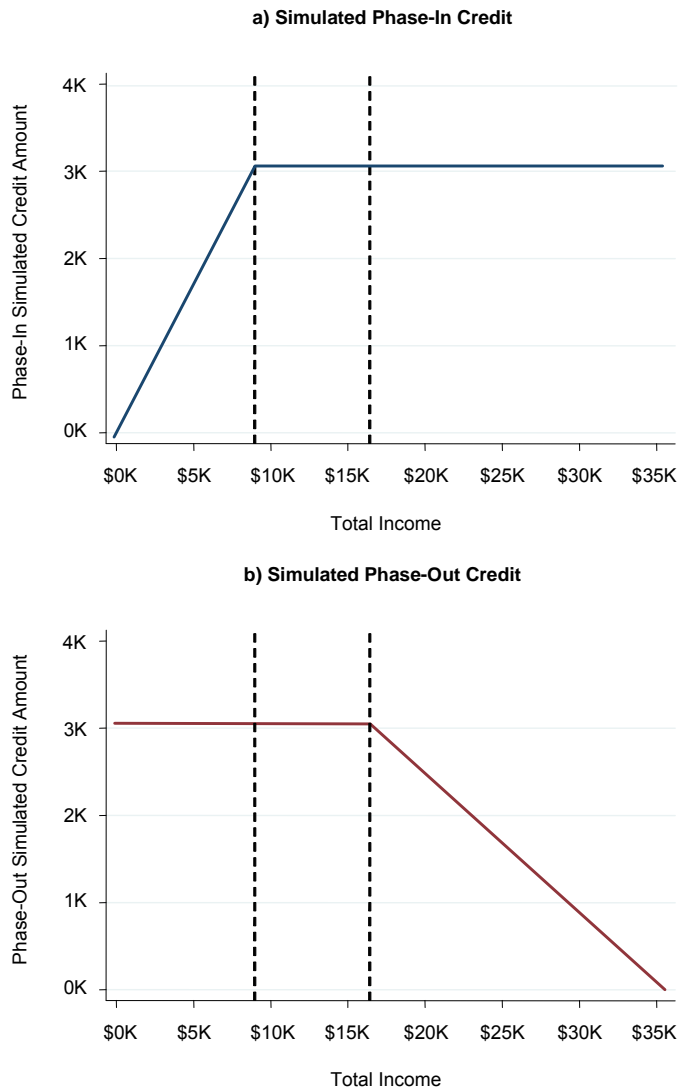
Notes: This figure plots the credit schedules used to identify phase-in and phase-out responses in Figure 15 and Table III. For the phase-in schedule, the simulated credit increases from $0 to $3,050 as income rises from $0 to $8,970 (corresponding to the actual EITC phase-in schedule). The schedule is then constant at $3,050 above $8,970 in wage earnings. For the phase-out schedule, the simulated credit is constant at $3,050 for incomes up to $16,690 and then decreases to $0 at a 15.98% rate (as does the actual EITC phase-out schedule).