

Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings: Web Appendix

Raj Chetty, John Friedman, and Emmanuel Saez

A Sample and Variable Definitions

This appendix expands on Section II.B to provide more detailed information on the construction of our analysis samples. As described in the text, we express all monetary variables in 2010 dollars, adjusting for inflation using the official IRS inflation parameters used to index the tax system.

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. *Total earnings* is the total amount of earnings used to calculate the EITC. This is essentially the sum of wage earnings and net self-employment earnings reported on the 1040 tax returns.¹ *Wage earnings* is the sum of wage earnings reported on all W-2 forms filed by employers on behalf of the primary and secondary filers. 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 1996-1998.² 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.³

We define the number of children as the number of children claimed for EITC purposes. The EITC children variable is capped at 2 from 1996-2008 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

¹More 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.

²Total 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.

³We have checked that our results are not driven by marriage effects by re-doing the analysis using solely individual earnings, instead of family earnings.

(e.g., because earnings are above the eligibility cutoff), we define the number of children as the number of dependent children.⁴

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 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 also remove observations with ITINs from the sample.⁵ 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 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 which an individual files a tax return, claims one or more children, and has income in the EITC-

⁴The 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.

⁵The 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.

eligible range.⁶ We then further restrict the data to individuals who move across 3-digit ZIP codes (ZIP-3s) in some year between 2000 and 2005. We impose this restriction to ensure that we have at least four years of data on earnings before and after the move. 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 and assign the individual to this first post-move neighborhood in all four subsequent years. 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.⁷ 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 individual 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 earnings as the total income reported on W-2 forms.⁸ We code total earnings as zero for non-filers who have no W-2’s.⁹ 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.¹⁰ 11.6% of

⁶We 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.

⁷As 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.

⁸Excluding 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.

⁹This 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.

¹⁰For individuals with multiple W-2 forms, we use the W-2 with the largest amount of earnings and non-missing

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). 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. 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. Over 97% of children are claimed as dependents on a tax return within 4 years of their birth. 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. 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.

B Supplementary Evidence

This appendix presents the details of robustness checks and additional tests that are described briefly in the text.

Impact of Other Taxes and Credits on Marginal Incentives

As described in the text, marginal tax rates in the income range we study are primarily determined by the EITC. This is illustrated in Appendix Figure 1, which shows that the schedule of total tax liabilities is very similar to the EITC schedule. There are two reasons for this. First, 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 also quantitatively small relative to the EITC; the maximum Child Tax Credit per child is \$500 before 2001 and \$1,000 starting in 2001. Second, federal income taxes and state income taxes typically affect the budget set starting in the phase-out region because of exemptions and deductions.

address information.

Furthermore, recall that 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.

Robustness to Alternative Bunching Measures

Appendix Figure 2 reproduces the main wage-earnings results (shown in Figure 7 and Appendix Figure 9a) using an alternative definition of sharp bunching. In this figure, 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 Appendix Figure 9a. To compare the coefficient in Panel A to that in Appendix Figure 9a, one must multiply the coefficient by 5.1 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 7. The coefficient in Panel B can be compared directly with the coefficient in Figure 7.

Diffusion of Sharp Bunching Over Time

We illustrate variation in sharp bunching (using the baseline definition in the main text) over time in Appendix Figure 3. To construct these maps, we divide the observations into deciles after pooling all years of the sample, so that the decile cut points remain fixed across years. In 1996, shortly after the EITC expanded to its current form, sharp bunching was prevalent in a few areas with a high density of EITC filers (southern Texas, NYC, and Miami). Bunching then spread throughout the country and continues to rise.

Total Earnings Distributions in the Lowest and Highest Bunching Deciles

Appendix Figure 4 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 top decile neighborhoods, 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, which is why we use these neighborhoods as a counterfactual for behavior in the absence of the EITC.

Movers Analysis: Supplementary Tests

While the sharp bunching presented in Figure 3a is perhaps the clearest evidence of response 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 Appendix Figure 5. Using a difference-in-differences specification analogous to that used in Figure 3a, 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 \$4,043 = \77 of this increase.¹¹ 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.

Note that the impact of moving on EITC refunds shrinks over time after the move. We speculate that this is due to a combination of factors: (1) a significant number of individuals move again within the next 4 years, so the differences in neighborhood at year +4 are smaller than at year 0, attenuating the differences over time. (2) Over a long time period, income growth may not be perfectly correlated for individuals who move to different types of neighborhoods. We focus on the sharp break from -1 to 0 in our main analysis for this reason.

To show the impact of moving on the earnings distribution more broadly, Appendix Figure 6 plots total earnings distributions in the years before and after the move for the three groups in Figure 3. This figure is constructed in the same way as Appendix Figure 4, 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.

Note that 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.

Relationship Between EITC Filer Density and Sharp Bunching

The correlation between density and sharp bunching suggests that agglomeration facilitates the diffusion of knowledge in dense areas. Appendix 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.

Tax Preparers and Sharp Bunching

Appendix 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

¹¹Roughly 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.

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).

Cross-Neighborhood Wage Earnings Comparisons: Supplementary Tests

In this subsection, we implement the three supplementary tests described in Section IV.A.

Earnings Impacts for Workers at Large Firms. As discussed in the text, 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 evaluate whether the differences in wage earnings distributions shown in Figure 5 appear for workers at large firms. Appendix Figure 8 plots the difference between the earnings distributions for the highest and lowest deciles from Figure 5. 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. The difference in earnings distributions between the highest and lowest sharp-bunching areas remains very similar when we restrict the sample to workers at firms with more than 100 employees.

One may be concerned that individuals in high-knowledge areas work in the formal sector up to the point where they maximize their EITC refund and then work in informal jobs. Two pieces of evidence suggest that this is unlikely. First, audit data show that the likelihood of misreporting total earnings is no higher for individuals who report wage earnings in the plateau. Second, 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.

Relationship Between Sharp Bunching and Wage Earnings Including All Neighborhoods. The analysis in Figure 5 and Appendix Figure 8 considers only the first and tenth deciles of b_{ct} , the areas with the least and most knowledge about the EITC schedule. In Appendix Figure 9a, 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.¹² Consistent with the earlier results, wage-earners in areas with high sharp bunching have earnings that produce significantly larger EITC refund amounts. Wage-earners in the highest-

¹²EITC 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 Appendix Figure 9a, we find very similar results within each of these demographic groups.

bunching areas earn EITC refunds that are on average \$122 (5.2%) higher than those living in the lowest-bunching (near-zero knowledge) neighborhoods. On average, a one percentage point increase in b_{ct} raises the EITC refund by \$15.9. As in Figure 8a, the relationship is concave: the marginal effects of higher sharp bunching lead to larger effects on wage earnings at lower bunching levels.

Movers Asymmetry Test for Wage Earners. Appendix Figure 9b 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 3b, restricting the sample to wage earners. Note that Appendix Figure 9b can be interpreted a first-differenced version of Appendix Figure 9a, relating changes in EITC refunds to changes in local knowledge for movers using our movers analysis sample. Appendix Figure 9b 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 opposite sign.¹³ We reject the null hypothesis that there is no kink in the slope of the control functions at 0 with $p < 0.0001$, echoing the pattern of learning and memory documented for the self-employed in Figure 3b.

Appendix References

Kleven, Henrik, Claus Kreiner, and Emmanuel Saez. 2009. “Why Can Modern Governments Tax So Much? An Agency Model of Firms as Fiscal Intermediaries.” National Bureau of Economic Research Working Paper 15218.

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.

¹³The 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.

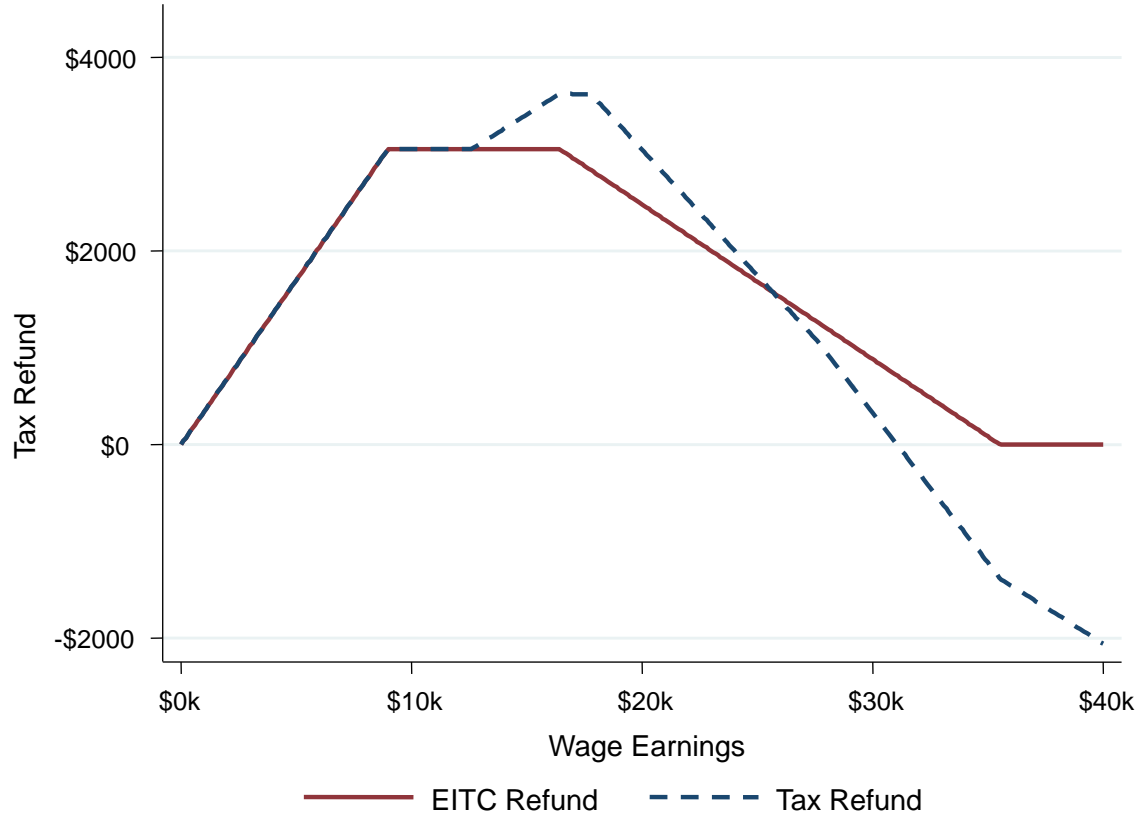
APPENDIX TABLE 1
Summary Statistics for Cross-Sectional Analysis Sample, 1999-2009

Variable	Mean (1)	Std. Dev. (2)
<u>Income Measures</u>		
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
<u>Tax Credits</u>		
EITC Refund Amount	\$2,543	\$1,454
Claimed EITC	88.9%	31.4%
Tax Professional Usage	69.6%	46.0%
<u>Demographics</u>		
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%
<u>Neighborhood (ZIP-3) Characteristics</u>		
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 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 (coded as 0 for states with no top-up).

APPENDIX FIGURE 1

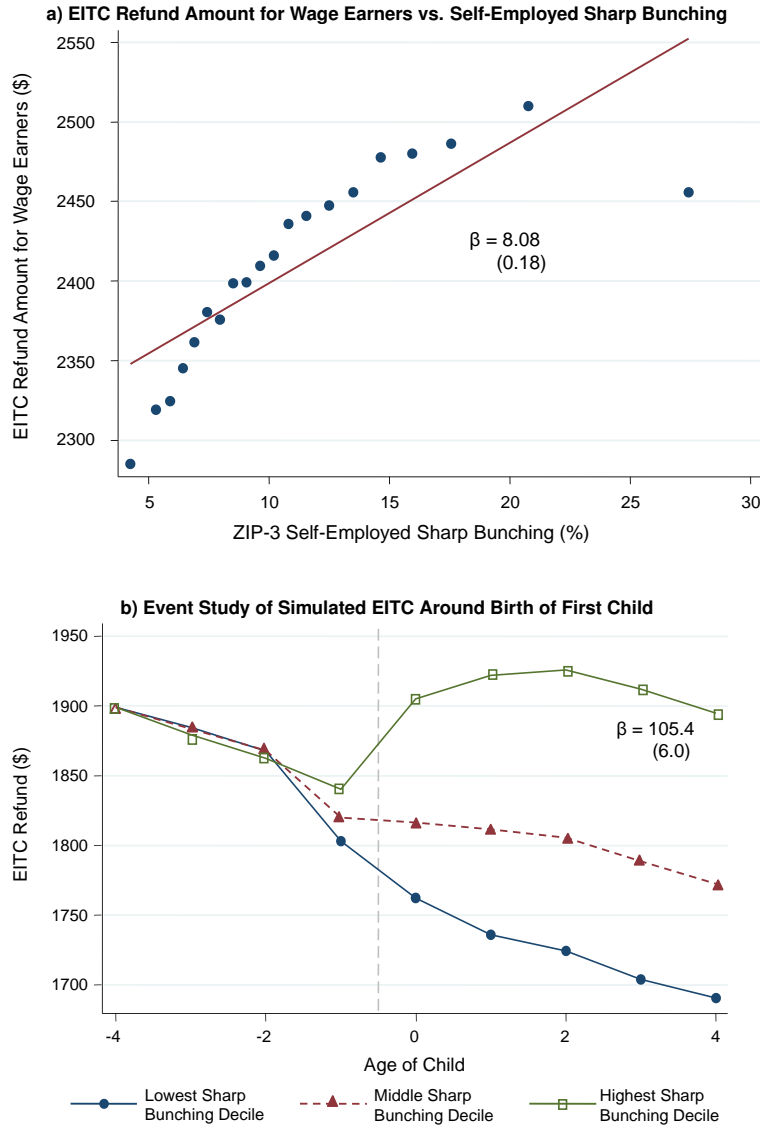
EITC Refund Schedule vs. Total Tax Liabilities for Single Filers with One Child



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.

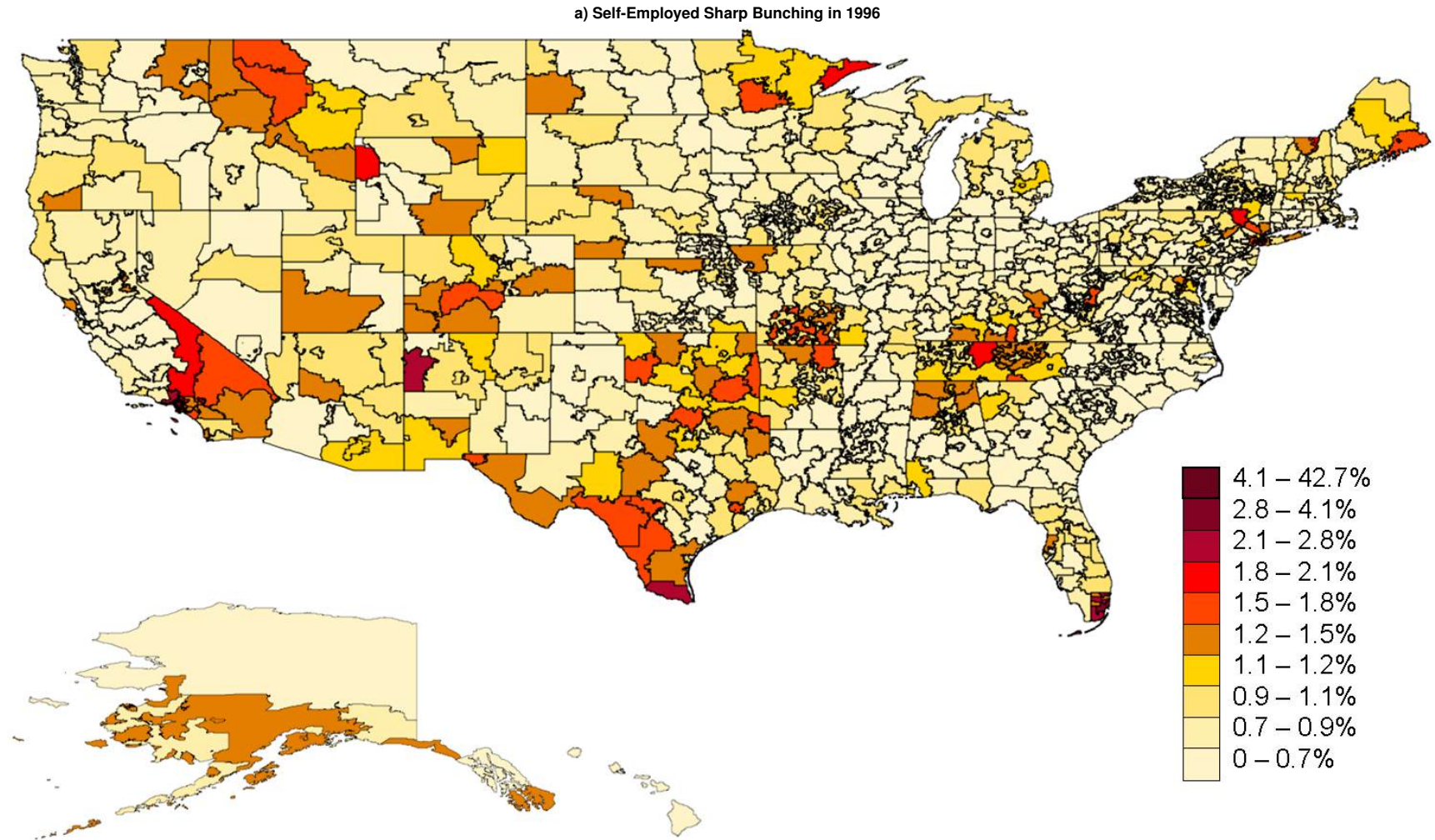
APPENDIX FIGURE 2

Results with Alternative Measure of Sharp Bunching



Notes: This figure reproduces the main wage-earnings results 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 Appendix Figure 9a. To compare the coefficient in Panel A to that in Appendix Figure 9a, one must multiply the coefficient by 5.1 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 7. The coefficient in Panel B can be compared directly with the coefficient in Figure 7.

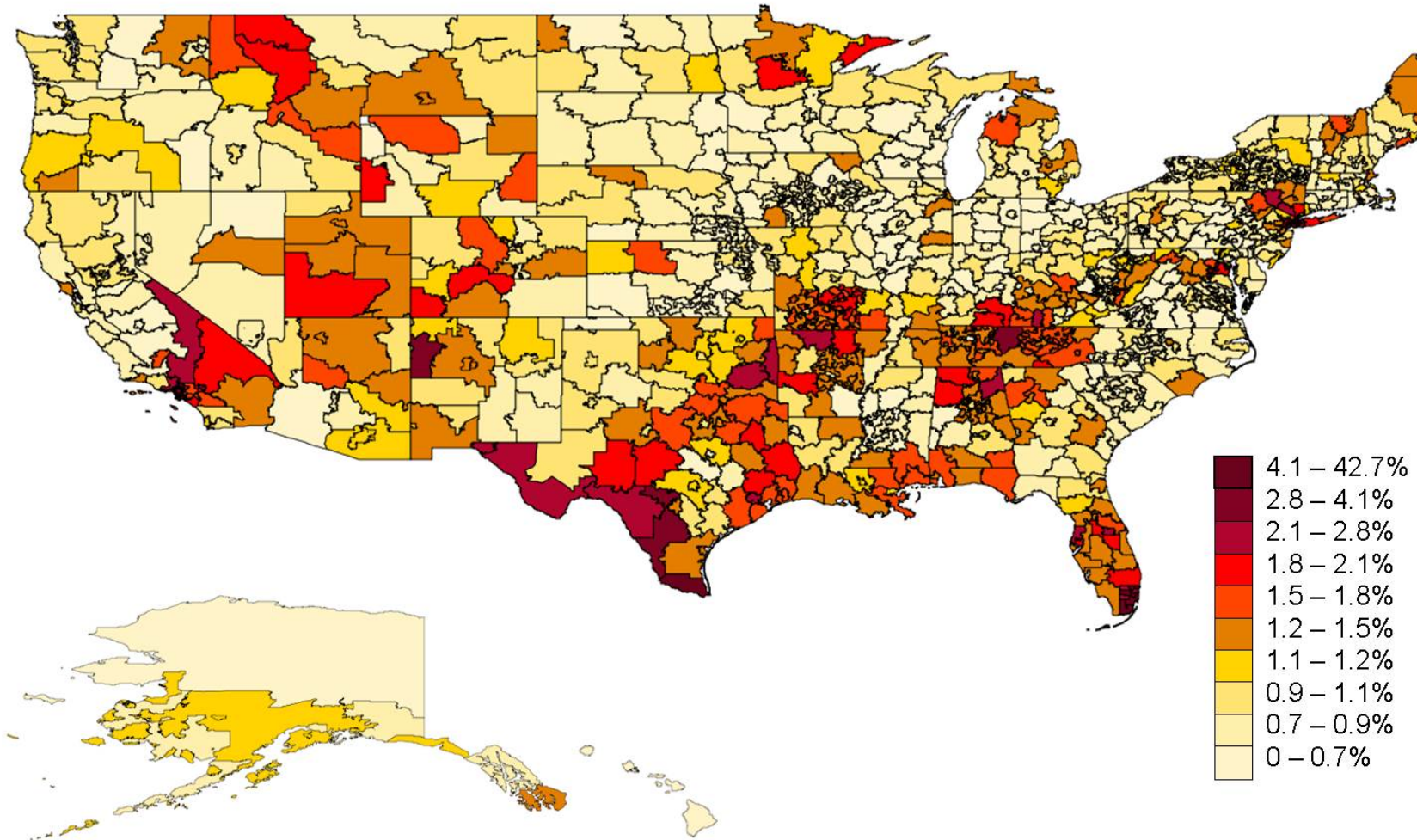
APPENDIX FIGURE 3
Self-Employed Sharp Bunching Rates Across Neighborhoods, 1996-2008



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.

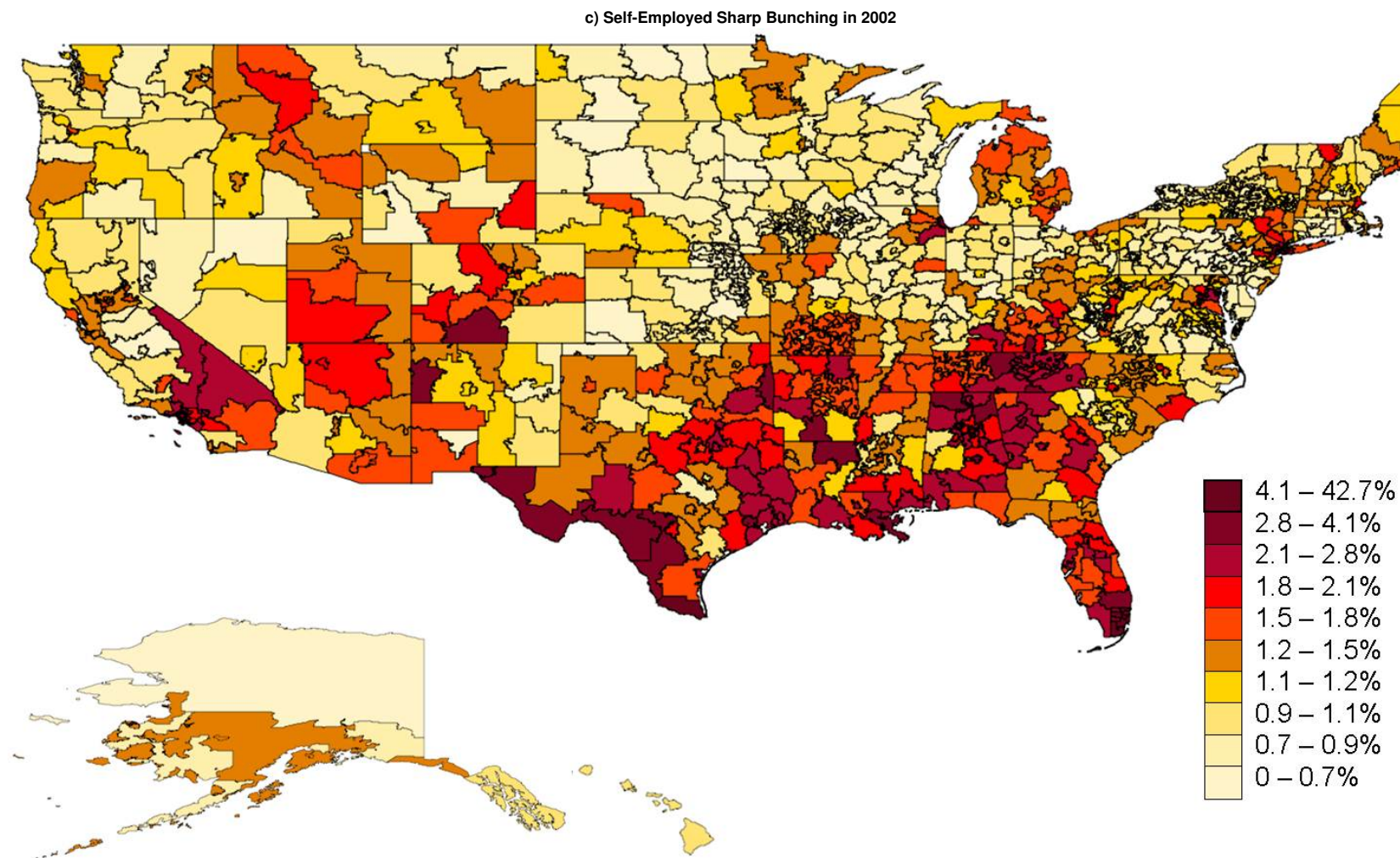
APPENDIX FIGURE 3
Self-Employed Sharp Bunching Rates Across Neighborhoods, 1996-2008

b) Self-Employed Sharp Bunching in 1999



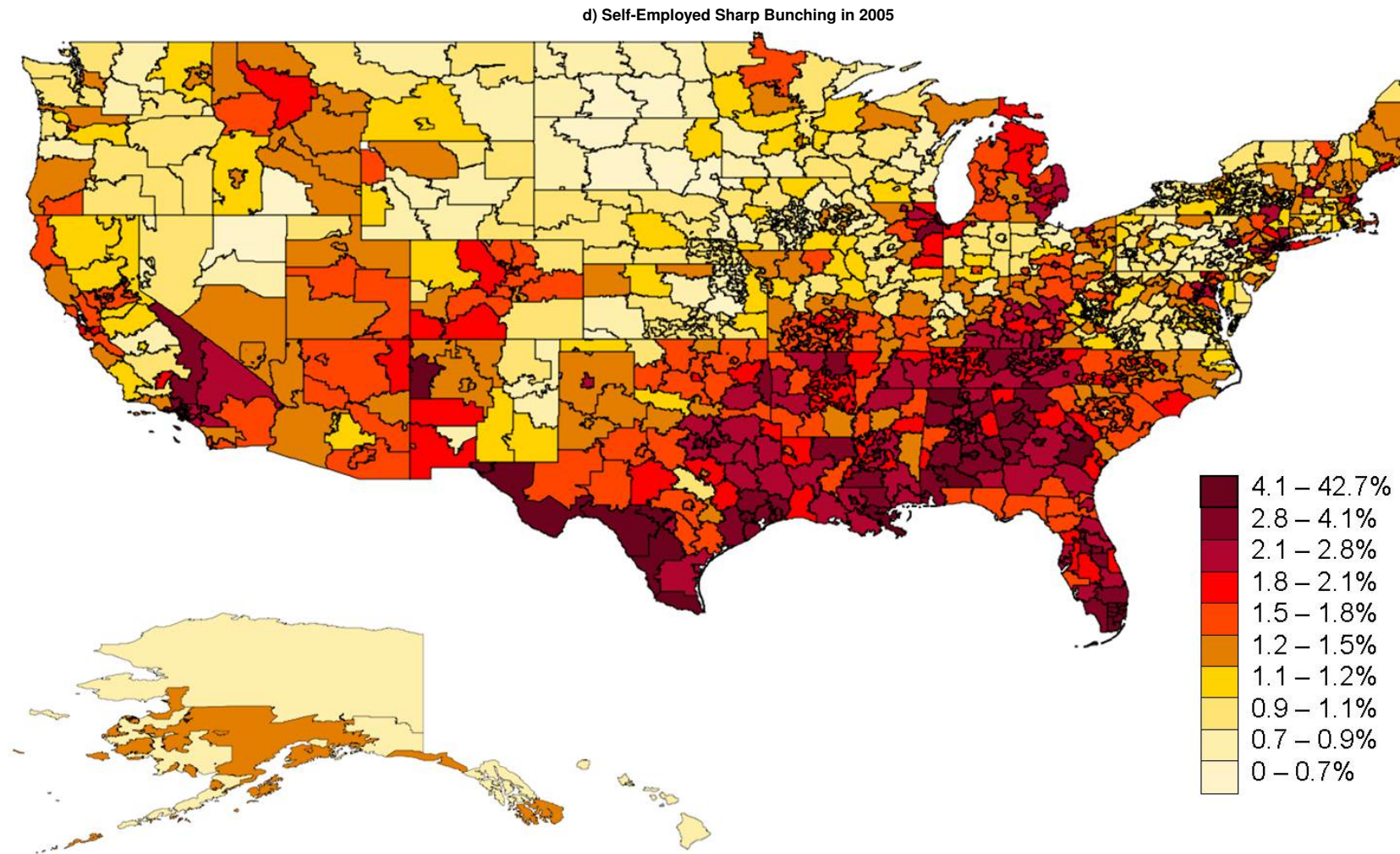
Notes: This figure replicates Panel A for the year 1999. Self-employed sharp bunching is defined as the fraction of all EITC-eligible households with children in the cross-sectional sample whose total income falls within \$500 of the first kink point and who have non-zero self-employment income. We divide 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



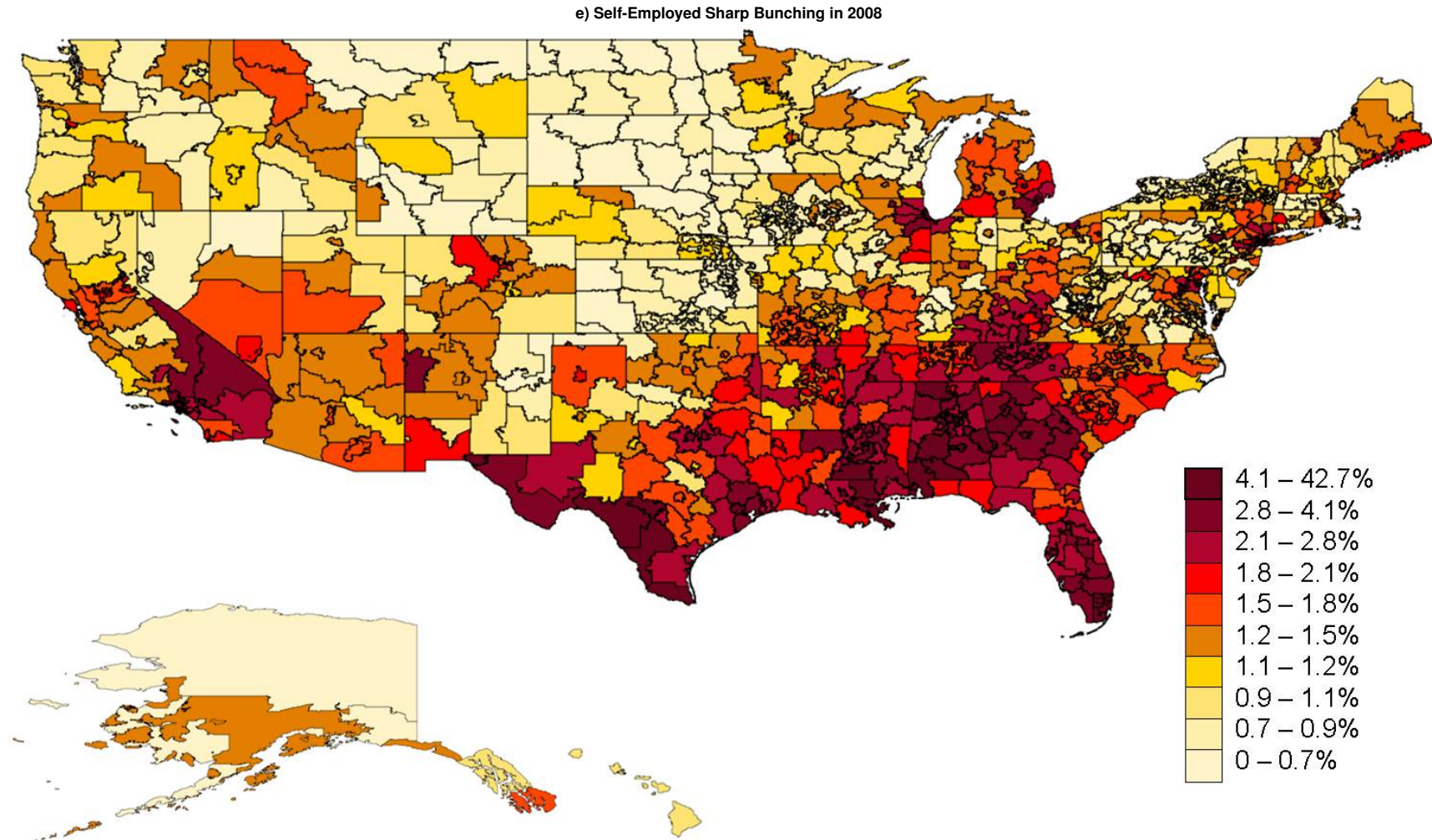
Notes: This figure replicates Panel A for the year 2002. Self-employed sharp bunching is defined as the fraction of all EITC-eligible households with children in the cross-sectional sample whose total income falls within \$500 of the first kink point and who have non-zero self-employment income. We divide 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



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.

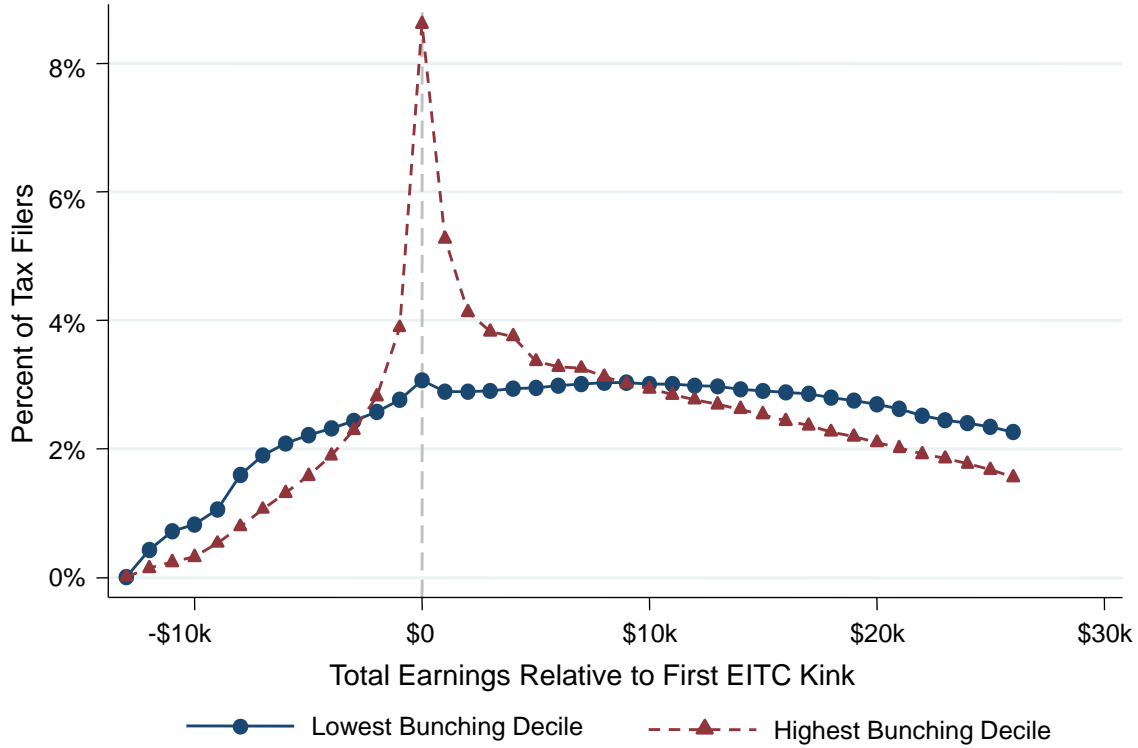
APPENDIX FIGURE 3
Self-Employed Sharp Bunching Rates Across Neighborhoods, 1996-2008



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

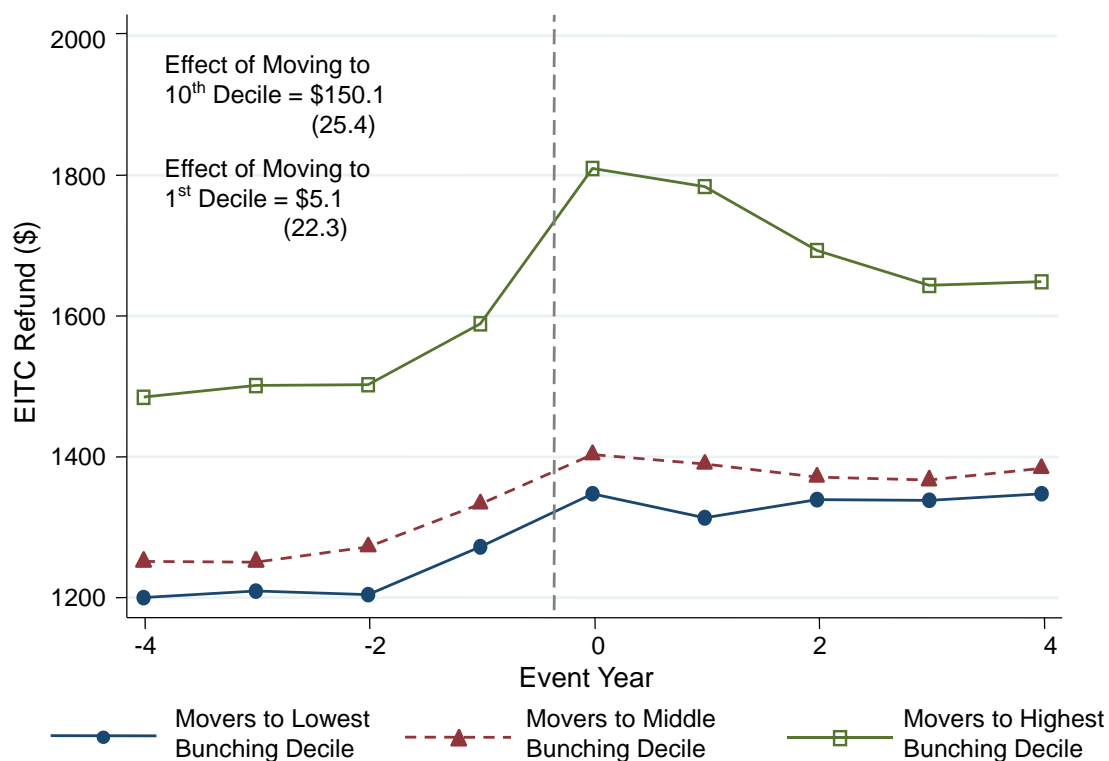
Earnings Distributions in Lowest vs. Highest Sharp Bunching Deciles



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.

APPENDIX FIGURE 5

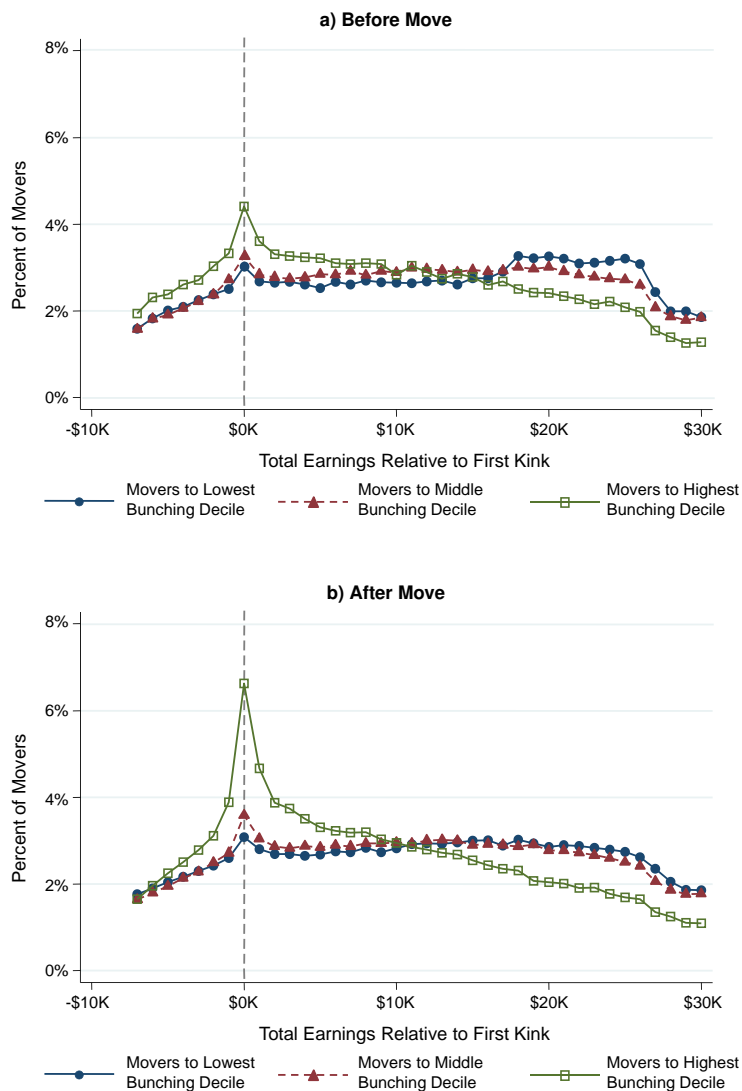
Event Study of Movers: Change in EITC Refund Amount



Notes: This figure plots an event study of the mean EITC refund amount among individuals who move from ZIP-3-by-year cells in the 5th decile to cells in the 1st, 5th, and 10th deciles. This figure replicates Figure 3a, with mean EITC refund amounts as the outcome instead of the rate of self-employed sharp bunching. See the notes to Figure 3 for details on sample and decile definitions.

APPENDIX FIGURE 6

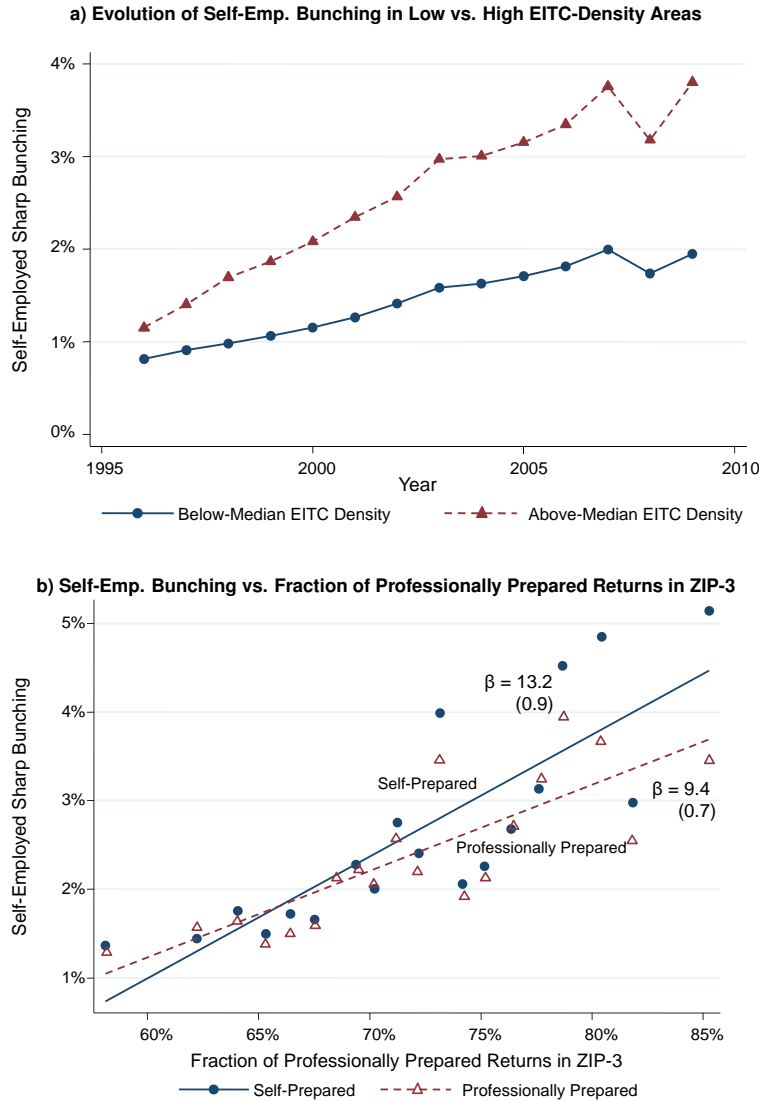
Total Earnings Distributions Before and After Move



Notes: These figures plot the distribution of total earnings before and after moving for the three groups of movers shown in Figure 3a. 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. 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 3 for details on sample and decile definitions.

APPENDIX FIGURE 7

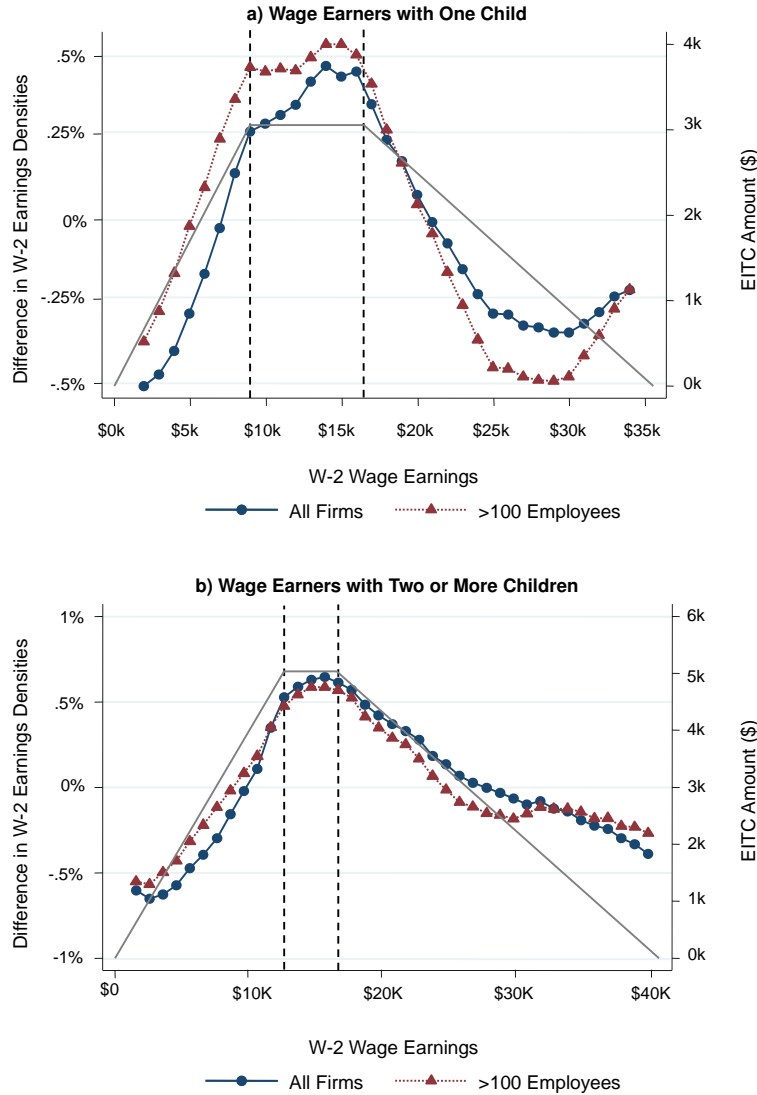
Correlates of Sharp Bunching



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.

APPENDIX FIGURE 8

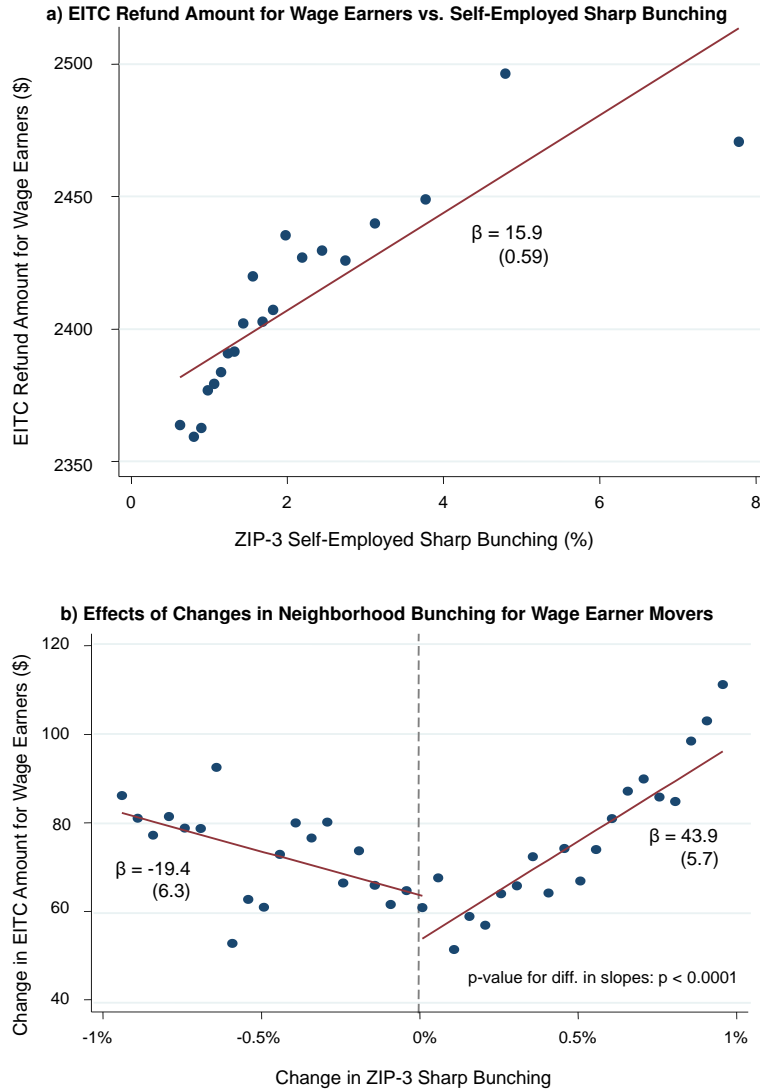
Differences in Wage Earnings Distributions: Lowest vs. Highest Bunching Deciles



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 5a; analogously, the series in circles in Panel B is the difference between the two series plotted in Figure 5b. 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 5 for further details.

APPENDIX FIGURE 9

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



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 3b, restricting the sample to wage earners and calculating the EITC refund based on W-2 wage earnings. See the notes to Figure 3 for more details on the construction of Panel B.