CARRA Working Paper Series

Working Paper *#2016-02*

The Effect of Low-Income Housing on Neighborhood Mobility: Evidence from Linked Micro-Data

Quentin Brummet
U.S. Census Bureau

Otávio Bartalotti
Iowa State University

Center for Administrative Records Research and Applications
U.S. Census Bureau
Washington, D.C. 20233

Paper Issued: May 16, 2016

The Effect of Low-Income Housing on Neighborhood Mobility: Evidence from Linked Micro-Data[+]

Otávio Bartalotti and Quentin Brummet[*]

May 13, 2016

Abstract:
While subsidized low-income housing construction provides affordable living conditions for poor households, many observers worry that building low-income housing in poor communities induces individuals to move to poor neighborhoods. We examine this issue using detailed, nationally representative microdata constructed from linked decennial censuses. Our analysis exploits exogenous variation in low-income housing supply induced by program eligibility rules for Low-Income Housing Tax Credits to estimate the effect of subsidized housing on neighborhood mobility patterns. The results indicate little evidence to suggest a causal effect of additional low-income housing construction on the characteristics of neighborhoods to which households move. This result is true for households across the income distribution, and supports the hypothesis that subsidized housing provides affordable living conditions without encouraging households to move to less-affluent neighborhoods than they would have otherwise.

Keywords: Subsidized Housing, Neighborhood Mobility, Low-Income Housing Tax Credits, Regression Discontinuity

JEL Classifications: H20, H31, I32, R23

[+] We are grateful to Matthew Freedman, Michael Hollar, Maggie Jones, Ann Owens, and seminar participants at American University, the Association for Public Policy Analysis and Management meetings, the Allied Social Science Associations SGE meetings, and the Population Association of America meetings for helpful comments. Thanks also go to Michael Hollar, Matthew Freedman, and Emily Owens for assistance with the construction of QCT classification measures. This paper is released to inform interested parties of ongoing research and to encourage discussion. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau. Any errors and omissions are our own.
[*] Bartalotti: Department of Economics, Iowa State University, 260 Heady Hall, Ames, IA 50011. Email: bartalot@iastate.edu. Brummet: Center for Administrative Records Research and Applications, U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233. Email: quentin.o.brummet@census.gov.

1 Introduction

Subsidized housing represents one of the largest transfers to low-income households in the United States. In 2014, the Department of Housing and Urban Development (HUD) spent roughly $50 billion on housing assistance for the poor (Congressional Budget Office 2015). Given that by definition housing assistance encourages households to move, a potentially important impact of subsidized low-income housing is its effect on the neighborhoods in which individuals choose to reside. Nonetheless, little evidence exists on whether subsidized housing has a causal effect on the location decision of households.

In large part, this lack of evidence is due to the absence of large-scale data on the origins of individuals moving into neighborhoods with subsidized housing. It is rare to find instances of credible identification strategies paired with rich data on the movements of new residents in neighborhoods. To overcome this hurdle, we use a large nationally representative panel data set derived from the 2000 and 2010 decennial censuses of the US population. We analyze the Low Income Housing Tax Credits (LIHTC) program, and use discontinuities in rules that determine program generosity to estimate the causal effect of LIHTC-induced construction on neighborhood mobility patterns. Specifically, our analysis makes use of three separate discontinuities in the designation process for Qualified Census Tracts (QCTs), which allocate additional funding for low-income housing development in specific geographic areas. Taken together, these rules provide exogenous variation in the amount of LIHTC-supported construction that can be used to assess the effect of low-income housing on residential mobility in a regression discontinuity (RD) strategy. Given that LIHTCs provided funding to roughly one third of new multifamily construction built in the US over the past thirty years, understanding whether LIHTC alters households' location decisions is extremely important for both the research and policy communities (Khadduri et al. 2012).

As a descriptive result, mobility patterns differ dramatically across areas based on the amount of LIHTC-subsidized housing. New residents in areas with high amounts of LIHTC construction moved to much less affluent neighborhoods when compared to new residents in areas with low amounts of LIHTC construction. However, our RD results provide little evidence to support the hypothesis that these descriptive differences represent causal effects of subsidized low-income housing construction on neighborhood mobility patterns. In particular, we find that new residents in neighborhoods exogenously shocked with increases in subsidized housing have moved from similar neighborhoods to their counterparts in counterfactual neighborhoods that did not receive the exogenous increase in new housing construction. These results are robust across a wide range of RD strategies and specifications, and hold for all levels of the household income distribution.

In addition to results that focus on the mobility patterns of new residents, referred to as "in-mobility," we also investigate whether low-income housing availability changes neighborhood composition. Consistent with previous studies, we find little effect of LIHTC construction on neighborhood characteristics such as poverty rate and median income. Hence, our results support the hypothesis that new residents in neighborhoods shocked by new low-income housing construction moved to similar neighborhoods than they would have in the absence of the low-income housing construction.

These results are relevant to the current policy debate around low-income housing in the United States. Proponents of subsidized housing point out that it provides affordable housing that is accessible to low-income households, but many critics worry that subsidized housing in low-income communities merely serves to induce poor households to move to poorer neighborhoods than they would have otherwise. While a large literature has examined these issues, we are the first to causally examine whether subsidized housing encourages individuals to move into poor neighborhoods on such a large scale. Given that recent evidence such as Chetty and Hendren (2015) highlights the importance of neighborhoods in determining social mobility, it is essential to understand the mobility effects of place-based policies such as LIHTC.

The rest of the paper is structured as follows. Section 2 provides information about the institutional details and previous literature studying the LIHTC program. Section 3 then describes the detailed microdata utilized in the analysis, while Section 4 presents results of analyses examining the effect of QCT status on LIHTC development, neighborhood mobility patterns, and neighborhood characteristics. Section 5 then concludes and discusses the implications of these results.

2 Background

The LIHTC program was created by the Tax Relief Act of 1986 as an attempt to increase the availability of low-income housing in the United States. To be eligible for these credits, developers must set aside a given percentage of a newly built or rehabilitated housing project to be rented to low-income individuals at HUD-determined rates.[1] If they qualify, developers are provided with a ten-year stream of credits.[2] Developments are required to be in place for 30 years, although the IRS will not seek to recapture tax credits after developments have been in place for 15 years, so developments sometimes lapse from low-income status after 15 years (Khadduri et al. 2012). In practice, these developments often serve very disadvantaged populations. Over 40 percent of the LIHTC units studied by Horn and O'Regan (2013) housed households that classified as extremely low income, and the majority of these households are receiving some other form of rental assistance.

Importantly, the amount of reimbursement from LIHTC differs based on whether the development is located in a QCT.[3] Developments in QCTs are entitled to up to a 30 percent higher tax credit, and QCT status is based on three criteria.[4] Specifically, tracts must pass one of two criteria to become eligible. First, a tract is eligible if at least half of its households fall below 60 percent of Area Median Gross Income (AMGI).[5] We will refer to this as the "eligibility" criteria due to the relationship of the 60 percent cutoff and HUD program eligibility rules. Second, a tract can be qualified if it has a

[1] Developers must either set aside at least 20 percent of units to be rented to individuals with incomes less than 50 percent of area median gross income (AMGI) or at least 40 percent of units for individuals with incomes of less than 60 percent of AMGI. Note that the size of the tax credit differs based on how many units are designated to be low-income units.
[2] While the size of the credit varies, a typical credit might reimburse 3.5 percent of building purchase costs and 9 percent of rehabilitation costs every year for 10 years.
[3] In this paper, we will focus our discussion and analysis on the form of tax credits that cover 70 percent of qualifying costs, and not cover the less widely used 30 percent credits that are allocated in a non-competitive fashion. Refer to Eriksen (2009) or Khadduri et al. (2012) for more thorough discussions of the LIHTC program.
[4] Much of this discussion of QCT status follows closely from Hollar and Usowski (2007).
[5] Throughout this paper, area refers to HUD-defined areas, which are typically Metropolitan Statistical Areas (MSAs), but are sometimes modified to account for substantial differences in housing markets within MSAs.

poverty rate of 25 percent or higher.[6] In addition to these two criteria, the total population of all QCTs in an area cannot exceed 20 percent of total area population. Therefore, in some areas eligible tracts are disqualified due to not falling in this 20 percent window. While selection for LIHTC relies in some part on subjective decisions by states,[7] QCT designations are not influenced by subjective guidelines.

A substantial previous literature has examined effects of subsidized housing on a number of important outcomes, but there is relatively little evidence on the effects of this subsidization on neighborhood mobility patterns.[8] For example, while previous literature on LIHTC shows that the increased low-income construction crowds out private housing investment, the crowd out is not one to one; rather, LIHTC is associated with at least modest increases in low-income housing in neighborhoods (Sinai and Waldfogel 2005, Baum-Snow and Marion 2009).[9] Moreover, a number of studies provide evidence that LIHTC investment increases housing prices in neighborhoods, likely due to the rehabilitation of older buildings.[10] In addition, Horn and O'Regan (2011) show that for a sample of three states, residents in LIHTC units are more likely to be minority than residents in surrounding census tracts.

Most similar to the current study are a group of studies that make use of exogenous variation induced by QCT designation rules to investigate the effects of LIHTC development. For example, recent work by Jackson and Kiwano (2015) finds evidence that LIHTC development reduces rates of homelessness. Most relevant to our analysis, Baum-Snow and Marion (2009) find that LIHTC developments lead to increases in the poverty rate in particular census tracts, while the results in Freedman and McGavock (2015) are mixed. These results are informative to the current study, but they speak to whether neighborhood composition changes, and not necessarily whether individuals are drawn

---

[6]Prior to 2002, there was no poverty rate criterion. Given that the current analysis studies projects put in place after 2002, these changes in QCT designation rules are not relevant to the current study.
[7] Specifically, states are required to file Qualified Action Plans (QAPs) that describe preferences for allocations.
[8] For a recent survey of work in this area, refer to Collinson et al. (2015).
[9] Note that even in the absence of crowd out, subsidized housing construction may change the relative price of rental housing and thereby create changes in neighborhood mobility patterns. For example, recent research on housing vouchers by Eriksen and Ross (2015) suggests that vouchers do not affect the price of rental housing overall, but do change the rental prices of specific housing units.
[10] Lee et al. (1999) and Green et al. (2002) provide descriptive evidence on the relationship between housing prices and LIHTC investment. See Schwartz et al. (2006) or Diamond and McQuade (2015) for more detailed causal analyses of the effect of LIHTCs on neighborhood housing prices.

into neighborhoods with LIHTC construction. Therefore, we consider the separate question of whether LIHTC affects mobility patterns between neighborhoods.[11]

3 Data

Our analysis uses linked person-level records from the 2000 and 2010 censuses of the US population. The sample is composed of all individuals in 2010 who appear in the 2000 Census long form data set.[12] To this data, we add information on LIHTC developments that is publicly available from HUD.[13] Last, we take publicly available data on tract characteristics from the long form of Census 2000 and 2008-2013 American Community Survey (ACS) 5-year estimates. These tract-level measures are used to construct both QCT eligibility rules that form the basis of our running variable and neighborhood characteristics that are used to construct our outcome variables.[14] Because we focus on neighborhood mobility, our analysis sample is composed of all individuals who moved census tracts between 2000 and 2010.

Table 1 displays summary statistics from the analysis sample broken apart by QCT status.[15] Panel A presents descriptive statistics of census tracts in year 2000. Given that QCT designations target poor census tracts by design, it is unsurprising to see that QCT tracts appear more disadvantaged along a number of dimensions. In particular, QCT tracts have much higher concentrations of minorities and people living in poverty. These tracts have over 30 percentage points lower fraction of white residents and poverty rates are on average over 20 percentage points higher than non-QCT tracts. QCT tracts are also much more heavily composed of renters. While non-QCT tracts have over 66 percent of their units occupied by owners, only 35 percent of housing units in QCT tracts are owner occupied. Last, note that

[11] Our analysis also relates to work such as Carlson et al. (2012), who examine the effects of housing vouchers on neighborhood mobility. Our analysis differs from this previous work in that it uses a nationwide sample and utilizes discontinuities in program rules rather than variation over time in order to identify the effects of subsidized housing.
[12] Section B.1 of the Data Appendix (Appendix B) provides a detailed description of the data set creation. The 2000 and 2010 censuses are linked together using a probabilistic linkage process that utilizes name, date of birth, and geographic location (Wagner and Layne 2013). A detailed discussion of this linkage process may be found in Section B.2 of the Data Appendix.
[13] These data are publicly available on the HUD website at http://lihtc.huduser.gov/.
[14] For a small number (less than 2 percent) of census tracts, we are unable to correctly assign QCT status. These tracts are dropped from the primary analysis presented below.
[15] Throughout the analysis, QCT status is constructed using 2000 Long Form Census data. These designations were used for years 2003-2006.

QCT tracts have slightly higher fractions of new residents than non-QCT tracts, but this difference is relatively minor compared to the differences in racial/ethnic composition and poverty rates. The question of whether LIHTC causes increases in the number of new residents will be discussed later, as it influences the interpretation of our main results pertaining to neighborhood mobility.[16]

Panel B of Table 1 presents statistics on neighborhood mobility by QCT status. Here, neighborhood mobility refers to baseline differences in tract characteristics between an individual's origin and destination census tracts in year 2000. Therefore, this measure picks up gaps between tracts in year 2000, but not any changes in tracts over time. These results show that new residents in non-QCT tracts have moved to a neighborhood that is on average very similar to their old neighborhood, or slightly more affluent. However, new residents in QCT tracts have recently moved to much different tracts from their previous tracts. In particular, new residents in QCT tracts live in a tract that was over 16 percentage points more renter heavy and more than 11 percentage points poorer than their previous tract in year 2000. Similar differences exist for racial composition and neighborhood income as well, and demonstrate that new residents in QCT tracts are much different from new residents in non-QCT tracts: on average, they have recently moved to much less affluent neighborhoods compared to new residents in non-QCT tracts. While these differences are practically very large, they are predictable since QCT status is assigned to poor tracts by construction.

While QCT tracts are much different from non-QCT tracts in terms of mobility patterns and baseline characteristics, Panel C of Table 1 shows that these two groups of census tracts evolve similarly in terms of neighborhood characteristics over the decade. In fact, QCT tracts became more heavily concentrated with white residents, more heavily owner-occupied, and less poor over the decade relative to non-QCT tracts. As with the statistics in Panel B, this evidence is descriptive, motivating the use of our RD analysis below to evaluate the causal impact of low-income housing development on neighborhood mobility.

[16] This question is particularly important, as previous work such as Woo et al. (2014) finds evidence that LIHTC construction may lead to increases in housing turnover.

4 Analysis

Estimating the causal effect of low-income housing on neighborhood characteristics and mobility would be straightforward if LIHTC development was randomly assigned across neighborhoods. However, LIHTC development is clearly targeted at particular neighborhoods, and therefore we pay attention to the method in which "treatment" is assigned. In this analysis, we use an RD design that exploits exogenous variation generated by discontinuities in program rules that designate QCT status. In particular, our baseline results are estimated using reduced form (RF) specifications such as the following:

$$Y_i = f(Eligibility_i, Poverty_i, Ranking_i) + \tau * QCT_i + \varepsilon_i \qquad (1)$$

Here, $Eligibility_i$, $Poverty_i$, and $Ranking_i$ are the three running variables described above, and our main treatment variable, QCT_i, indicates whether a tract has crossed the treatment threshold and is designated as QCT. Note that these are sharp regression discontinuity estimates, as passing the threshold exactly predicts QCT status. ε_i is an error term, which is assumed to evolve continuously at the treatment threshold.[17]

Y_i represents the outcome of interest, which in our case will be measures of neighborhood composition or mobility patterns. To further understand these measures, consider an individual who moves from her sending neighborhood *i* in year 2000 to receiving neighborhood *j* in year 2010. The eventual measure of interest is the difference between neighborhood characteristics in 2010, $X_{j,2010}$, and an individual's neighborhood in year 2000, $X_{i,2000}$. For the purposes of our analysis, we separate this measure into two pieces. The first piece, which we refer to as neighborhood mobility, can be represented as $X_{j,2000} - X_{i,2000}$. This measure captures differences in neighborhood characteristics between sending and receiving neighborhoods at the start of the decade, and allows us to estimate if an individual's location decision was shaped by low-income housing development. There may be other effects of low-income housing, though. Consider the case of average neighborhood income. If high-income individuals

[17] Refer to Papay et al. (2011) for a more detailed discussion of RD designs with multiple running variables.

move out in response to the low-income housing development or the additional low-income housing worsens job opportunities then this would be reflected in the average income of the tract. Therefore, we also consider measures of changes in neighborhood characteristics, $X_{j,2000} - X_{i,2000}$. This allows our analysis to estimate if LIHTC-supported construction affects the evolution of neighborhood characteristics between 2000 and 2010.

As mentioned in Section 2, the structure of QCT designation rules generates discontinuities in multiple running variables that provide exogenous variation in low-income housing construction for our analysis. To accommodate this complexity, we run regressions including a function of the three running variables, $f(Eligibility_i, Poverty_i, Ranking_i)$.[18] In addition, we choose a bandwidth to implement the RD design. For the purposes of our analyses, we choose preferred bandwidths of 0.40 for eligibility, 0.12 for poverty rate, and 0.15 for the population cap ranking. We utilize these bandwidths in our baseline specifications for two reasons. First, when using each running variable independently, these bandwidths roughly correspond to those chosen using the cluster-robust MSE-optimal bandwidth selection procedure in Bartalotti and Brummet (2016) for several key dependent variables.[19] In addition, these bandwidths produce relatively equal sample sizes when used separately. Nonetheless, because the bandwidths we choose are admittedly *ad hoc*, Section 4.3 will present results using a range of possible bandwidths as robustness checks.

Additionally, we present IV regressions that use the exogenous variation induced by QCT status to examine the effect of low-income housing units and projects induced by the extra LIHTC funds on neighborhood characteristics and mobility. We measure LIHTC development using the number of new low-income LIHTC projects.[20] This leads us to the following estimating equations:

$$Y_i = f(Eligibility_i, Poverty_i, Ranking_i) + \tau * LIHTC_i + e_i \qquad (2)$$

[18] In practice, our baseline specifications will parameterize $f(.)$ using a fully interacted quadratic function, and we perform robustness checks to assess the sensitivity of our results to this assumption.
[19] In practice, these bandwidths are slightly larger than those selected using the Bartalotti and Brummet (2016) procedure to account for loss of precision due to the use of multiple running variables.
[20] Results using number of low-income LIHTC units are qualitatively similar and available from the authors upon request.

Here, $LIHTC_i$ is a measure of low-income, LIHTC-subsidized housing development. In practice, we will measure LIHTC development using the number of projects built in tract *i*.[21]

The main assumptions underlying this identification strategy are that the endogenous variables measuring low-income housing supply are related with QCT status conditional on the running variables, and that any relevant unobserved factors that determine neighborhood characteristics are uncorrelated with the instrument conditional on the function *f(.)* near the cutoff. We will investigate the former assumption in Section 4.1, which presents first stage relationships between QCT status and LIHTC development. Given that these IV specifications are types of fuzzy RD designs, the second assumption can be restated as an RD assumption that the conditional expectation of Y_i evolves smoothly at the treatment threshold. In particular, we want to rule out systematic manipulation of the running variables. Given that QCT classification is based on data collected as part of the 2000 decennial census, it is doubtful that households or other institutions are able to systematically manipulate their tract's data in order to receive extra tax credits.[22] However, one may still worry that systematic measurement error in QCT classification is invalidating the assumption that the error term is continuous at the cutoff. To examine this assumption, we plot the marginal distribution of the three running variables in Figure 1. For all three running variables, there is very little evidence of a difference in the distribution of the running variable at the cutoff point. In Section 4.3, we further test this assumption and perform a number of robustness checks to attempt to measure the validity of this assumption.

4.1 The Effect of QCT Status on LIHTC Development

As a first step towards understanding the effect of LIHTC on neighborhood mobility, we investigate whether QCT status has a distinguishable effect on LIHTC development. As mentioned in Section 2, the QCT classification system leads to three separate running variables for the RD design: the "eligibility"

[21] Similar to Freedman and McGavock (2015), we restrict our attention to projects placed in service between 2004 and 2009.
[22] Refer to Lee and Lemieux (2010) for a discussion of the impacts of imperfect manipulation of the running variable on the interpretation of RD estimates.

running variable (i.e., the ratio of median household income to 60 percent of AMGI), the poverty rate, and the relative rank of the tract along these dimensions when compared to other tracts in the area. Therefore, as a starting point, we consider the variation in LIHTC construction induced by each of these variables separately.

Figure 2 presents the relationship between LIHTC development and the eligibility running variable. If this ratio is less than one, the tract passes the eligibility criterion for QCT status; if the ratio is greater than one, the tract fails the eligibility criterion. Panel A presents the relationship between this eligibility running variable and QCT status, whereas Panels B and C present the relationship between the running variable and the number of LIHTC units and projects, respectively. There is a pronounced discontinuity in QCT status, with tracts just to the left of the cutoff being over 40 percentage points more likely to receive QCT status. While Panels B and C do not rule out potential effects of QCT status on LIHTC development, the results are imprecise and unclear.

Figure 3 presents these same results for the poverty rate criterion. Recall that a tract must have a poverty rate of at least 25 percent to pass this criterion, implying that tracts to the right of the threshold are treated. These results show that this criterion binds quite dramatically, and produces pronounced increases in both the probability of receiving QCT status and LIHTC construction. Comparing Figure 3 to Figure 2, the discontinuities are slightly more pronounced for the poverty rate running variable, implying that the poverty rate criterion is more predictive of QCT status than the eligibility criterion. This relationship also shows up in Panels B and C, which present suggestive evidence that passing the poverty rate criterion has an effect on LIHTC development.

Last, Figure 4 presents these same graphs for the population cutoff criterion. Recall that only 20 percent of a given area can live in a QCT. To achieve this objective, tracts are ranked according to poverty rate and eligibility in order to keep only the top 20 percent of the area's population. Hence, we construct a running variable that is equal to the cumulative percentage of population that lives in a higher ranked track. Given that this criterion only binds when more than 20 percent of area residents live in tracts that qualify under either the eligibility or poverty rate criteria, it is unsurprising to see smaller discontinuities than in Figures 2 and 3. Nonetheless, every panel shows some suggestive evidence that this population cap criteria produces at least small changes in QCT status and LIHTC development at the threshold.

Table 2 displays estimates that summarize the results shown in the previous figures. Each coefficient represents the estimate of crossing the qualification threshold, where the regressions are based on specifications such as that shown in Equation (1) with only a single running variable. Note that all running variables have been redefined so that the estimated discontinuity captures the effect of obtaining QCT status. The results largely confirm the plots in Figures 2-4. In particular, specifications using all three running variables produce statistically significant discontinuity estimates for the increase in the probability of obtaining QCT status, but the estimate using the population cap ranking running variable is an order of magnitude smaller than the estimates using either the eligibility or poverty rate running variables. This difference carries over to first stage estimates of crossing particular thresholds on LIHTC construction. The poverty rate running variable specifications produce statistically significant effects of crossing the threshold on the number of LIHTC projects, whereas the specifications using the eligibility running variable produce practically important, but statistically insignificant effects. As might be expected, the specifications using the population cap running variable are small and too imprecise to produce meaningful conclusions. Panel A presents results using a regression including a fully interacted quadratic specification of all three running variables, which form the first stage for our IV estimates below. Here, we see that obtaining QCT status is associated with 11.21 more LIHTC units and an additional 0.12 LIHTC projects. Both of these estimates are statistically significant at the 0.01 level.

4.2 The Effect of LIHTC Development on Neighborhood Characteristics and Mobility

Prior to providing RD estimates based on the RF specification shown in Equation (1), Figures 5-7 present the graphical relationship between our outcomes of interest and the three running variables. In each of these figures, a single running variable is shown on the horizontal axis while the vertical axis contains average changes in tract characteristics for new residents. Figure 5 first presents this relationship for the eligibility criteria. As can be seen, there is little evidence to indicate that there is a discontinuity in mobility patterns at the cutoff for the majority of outcomes. In particular, new residents in QCT tracts do not appear to have been induced to move to neighborhoods that are less white or lower income than they would have in the absence of QCT status.[23] The two exceptions are poverty rate and fraction Hispanic, which display small discontinuities at the eligibility cutoff, leaving open potential consequences of LIHTC development on neighborhood mobility. Note, however, these results are sensitive to outliers and bandwidth choice.

Figure 6 presents the same results for the poverty rate criterion, which display no perceptible discontinuities across any of the various outcome variables. Given the strong relationship between crossing this threshold and LIHTC development shown in Figures 2 and 3, this suggests that there may be little effect of LIHTC development on neighborhood mobility patterns. Finally, Figure 7 shows results for the population cap running variable, which aligns with those using the poverty rate running variable – there is very little evidence that new residents are moving to less affluent tracts because of the QCT status.

Table 3 presents estimates of the discontinuities shown above in Figures 5-7.[24] Unsurprisingly, the results conform to the plots in Figures 5-7. Given the results shown above, Panels A and B are of particular interest because the eligibility and poverty rate thresholds have the largest effect on QCT status and LIHTC development. Examining these two variables separately, we see very small effects estimated for fraction of owner-occupied housing units and racial/ethnic composition in the census tract. While estimates using either running variable indicate that new residents have been induced to move to poorer neighborhoods than they would have otherwise, the results are not statistically significant for the eligibility running variable. While these results present mixed evidence on the effect of LIHTC on individual movement across neighborhoods, the results are sensitive to changes in bandwidth choice and

[23] It is important to consider racial compositions of tracts in the analysis, as previous descriptive work has found that racial composition is a predictor of LIHTC development (Rohe and Freeman 2001).

[24] Table A.1 of the Supplemental Results Appendix (Appendix A) contains a similar specification to that shown in Table 3 with a linear specification for the running variable. The results are qualitatively similar.

motivate the use of our preferred specifications below. These specifications, reported in Table 4, combine all running variables and are more robust to changes in bandwidth choice.

In order to pool these estimates together, Table 4 presents our main RF results based on the specification outlined in Equation (1). Panel A provides our baseline specifications, which test whether QCT status influences mobility patterns so that these neighborhoods receive more movers from less-affluent neighborhoods. Examining these results, we see very small and precise RF estimates of the effect of crossing the QCT threshold on neighborhood mobility patterns. Taken at face value these estimates indicate that crossing the QCT threshold causes neighborhoods to receive new residents from neighborhoods that are only 0.06 percentage points less renter heavy and 0.15 percentage points less heavily concentrated with white residents than their counterparts moving into non-QCT tracts. Neither of these estimates is statistically different from zero, and the 95 percent confidence intervals reflect reasonably precise estimates. Hence, these results indicate very little effect of the additional low-income housing construction on mobility of households into poor neighborhoods, supporting the hypothesis that these households did not move to poorer neighborhoods than they would have otherwise.

The main outcome variable in the analyses above is defined as the difference in baseline characteristics between census tracts in year 2000. Therefore, it is also important to know whether QCT status caused changes in neighborhood characteristics over the decade. To investigate this, we run specifications such as those shown in Equation (1), but replace the outcome variable with changes in tract characteristics over the decade for a given tract. Panel B of Table 4 presents these results, which demonstrate that there was little neighborhood change induced by the QCT designation across a number of different dimensions. The two exceptions are the estimates for changes in the fraction of African-American and Hispanic residents, which are marginally significant at the 0.10 level. Nonetheless, for the outcomes representing the economic composition of the census tracts we see no such evidence.

The final column of Panel B tests whether QCT status increases the fraction of new residents in a given census tract. If this were the case, one might be worried that the neighborhood mobility results in part reflected a changing share in the number of new residents in a given tract. The results do not support this hypothesis, however, as the estimate of the effect of QCT status on the fraction of new residents is statistically indistinguishable from zero and reasonably precise. Therefore, there is no evidence that changes in the size of new resident flows are driving the results shown in Panel A of Table 4.

To understand the magnitude of these estimates, Table 5 presents IV estimates based on specifications as described in Equation (2), which can be interpreted as the impact of additional LIHTC-subsidized projects in neighborhoods around the QCT designation cutoffs. Panel A presents the results for neighborhood mobility patterns, and we again see that there are no statistically significant effects of additional LIHTC construction on neighborhood mobility patterns. An additional LIHTC-subsidized project leads to very small changes in any of the outcome measures we study. These IV estimates are imprecise, but note that these developments are large relative to the size of a given census tract.[25] Therefore, the fact that the upper bounds of the 95 percent confidence intervals suggest a less than 10 percentage point effect on neighborhood mobility patterns rules out the hypothesis that the entirety of the descriptive in-mobility patterns shown in Table 1 were driven by LIHTC construction. Nonetheless, these may still be effects of LIHTC development on neighborhood composition even if mobility patterns are unchanged. To test this hypothesis, Panel B of Table 5 presents IV results for the effects of LIHTC development on neighborhood characteristics. Supporting the results in Table 4, none of the coefficients are statistically significant. Hence, in addition to there being very little evidence to suggest that LIHTC developments affect neighborhood mobility patterns, there are also relatively small effects of these developments on the evolution of neighborhood characteristics across the decade.

4.3 Robustness Checks

Given the importance of bandwidth selection in RD studies, it is important to examine the sensitivity of these results to bandwidth selection. In examine the robustness of our main results to bandwidth choice, Figure 8 presents the results of specifications such as those represented by Equation (1) where the

[25] For example, a typical LIHTC project consisting of 100 units would be sizeable relative to the average new resident population of tracts in our sample, which is roughly 300 individuals.

bandwidth has been altered.[26] Each panel in Figure 8 plots estimates and corresponding 95 percent confidence intervals for a separate dependent variable across different bandwidth selections.[27] The vertical line on each graph indicates the bandwidth that was used for the estimates in Table 3. Examining Figure 8, it is clear that these estimates are not driven by a particular bandwidth choice. While smaller sample sizes created by small bandwidths decrease precision, the majority of the evidence presented in Figure 8 supports the hypothesis that the main results presented in Table 3 are robust to changes in bandwidth.

Another concern is that these results somehow reflect pre-existing trends in tract characteristics that are not adequately controlled for by the running variables.[28] One method to check for such a problem is to run specifications such as those shown in Equations (1)-(2), but substitute baseline year 2000 tract characteristics as outcome measures in levels. This serves as a placebo test, because there should be no effect of LIHTC construction post-2000 on year 2000 tract characteristics.[29] Therefore, if our strategy is valid we would expect these estimated effects to be close to zero. Table 6 shows the results of this placebo test using tract-level aggregate data. The vast majority of these estimates are very small and statistically insignificant. The one exception is the statistically significant and practically important estimate for a positive relationship on pre-existing fractions of owner-occupied housing. Hence, it may

[26] Tables A.2 and A.3 in Appendix A present the same results as shown in Tables 3 and 4, but use a linear running variable specification. The results are qualitatively similar.

[27] Given that there are three running variables, one could conceivably vary the specification by altering any of the three bandwidths. To make the results presentable, the estimates in Figure 8 are obtained by increasing or decreasing each of the three bandwidths by a constant percentage. For example, the smallest bandwidths used in Figure 8 include bandwidth values for each of the three running variables that are one half of the value that was used for the estimates presented in Table 4.

[28] One may also worry that these results are significantly different in areas that are Difficult Development Areas (DDAs), given that DDAs are also eligible for a 30 percent increase in their tax credits. However, as shown in Appendix Table A.4, results restricted to non-DDA areas produce qualitatively similar estimates. DDA classification data is available online from HUD at https://www.huduser.gov/portal/datasets/qct.html.

[29] One exception might be that if there is strong persistence in the QCT qualification criteria in tracts over time, we might expect there to be an effect because the 2000-2010 QCT criteria are picking up effects of pre-2000 QCT construction. Therefore, this procedure serves as a useful placebo test.

be that our estimates are biased towards finding that new residents recently moved to more heavily owner-occupied neighborhoods.[30]

4.4 Heterogeneity in Results

Last, we consider potential forms of heterogeneity in these results. This is important, as the interpretation of our findings depends on whether these policies have differential effects across subgroups. First, because many observers often worry about the segregating effects of low-income housing, it is natural to examine whether these results differ across racial categories. In order to examine this question, Table 7 presents results estimated separately for white and black individuals.[31] Panel A presents results for white individuals, where the estimates are derived from IV specification such as that shown in Equation (2). These results are extremely imprecise, but the point estimates would suggest that white individuals were encouraged to move to neighborhoods that were more heavily African-American than they would have otherwise. Therefore, the results suggest that the LIHTC construction had an integrating effect by encouraging white individuals to move to less-white neighborhoods than they would have otherwise. Nonetheless, given the imprecision of the results, we lack substantive evidence to the question of whether LIHTC construction has differential impacts on neighborhood mobility across white and African-American individuals.

Given the concern over "poverty concentration" that surrounds many critiques of low-income housing, it is natural to also examine differential effects across income categories. If LIHTC housing indeed concentrates poverty, we might expect that the effect of LIHTC construction on neighborhood mobility would vary on the basis of household income. To check this hypothesis, Table 8 estimates our main IV specification outlined in Equation (2) across quartiles of the household income distribution in year 2000. Each cell corresponds to an estimate from a separate regression, and the panels of the table

[30] Appendix Table A.5 presents these same results using a linear running variable. These results are qualitatively similar to those shown in Table 6.
[31] These racial groups do not include individuals who identify as Hispanic or Latino, who are removed from this analysis. Results for Hispanic individuals are available from the authors upon request.

represent samples that are broken apart by quartile of the household income data in the entire sample. These results are consistent across the entire income distribution, as all quartiles show very small estimated effects. While the results are again somewhat imprecise, they do suggest that our main results presented in Section 5.2 do not appear to mask significant heterogeneity in effects across the income distribution. These results are also of immediate relevance to policy makers, as it documents that LIHTC-supported construction does not induce low-income households to move into lower-income communities than they would have otherwise.

There are a number of other dimensions along which the results might vary. First, different age groups might respond differently to the expanded low-income housing options given their different propensities to move. While not shown here, analyses broken apart by age of individual do not point to differential effects.[32] In addition, while it is natural to think that owners and renters might respond differentially to the availability of low-income housing, our results appear similar when the sample is split by whether or not the household owned a home in year 2000.[33] Last, previous studies such as Baum-Snow and Marion (2009) have shown that tract characteristics may change differentially based on whether the neighborhood was gentrifying between 2000 and 2010. There is no evidence of this effect in the current setting as our baseline neighborhood mobility results appear similar when broken apart by whether or not the tract gentrified over the decade.[34]

6 Conclusion

Subsidized low-income housing development creates housing opportunities for poor households, potentially creating significant effects on household location decisions. Because of this, many critics worry that these polices encourage poor individuals to move to low-income areas. Nonetheless, evidence on this question is lacking. We present some of the first causal large-scale evidence on the effects of targeted subsidized housing policy on neighborhood mobility by examining the effects of LIHTC

[32] Refer to Appendix Table A.6 for these results.
[33] Refer to Appendix Table A.7 for these results.
[34] See Appendix Table A.8 for the results of this analysis.

development on patterns of movement of individuals between neighborhoods. Using program rules for designating QCTs as a source of exogenous variation, our results suggest that the additional low-income housing does little to change either the way in which individuals move across neighborhoods or the composition of the neighborhoods themselves. In analyzing potential heterogeneity in the effects of low-income housing developments, we find plausible but imprecise heterogeneity by race and no difference across the income distribution in how LIHTC construction affects mobility patterns. This is important to policy makers, as it implies that LIHTC construction does not draw poor households to low-income neighborhoods.

While a number of previous studies examine direct effects of living in subsidized housing developments on a variety of outcomes, few test whether these developments affect household location decisions.[35] The results presented here provide evidence on the effects of these policies on a separate outcome: neighborhood mobility. We provide evidence that across all income levels, additional low-income housing construction does not incentivize individuals to move to less-affluent neighborhoods than they would have otherwise. Given the potential importance of neighborhood effects, these results lend key evidence to the debate over the effectiveness of subsidized low-income housing.

[35] For example, see Olson et al. (2005), Susin (2005), Jacob and Ludwig (2012), Jacob, Ludwig, and Miller (2013), or Jacob, Kasputin, and Ludwig (2015) for evidence related to public housing or housing vouchers.

References

Bartalotti, Otávio and Quentin Brummet. 2016. "Regression Discontinuity Designs with Clustered Data: Variance and Bandwidth Choice." In Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, volume 38), ed. M. D. Cattaneo and J. C. Escanciano. Emerald Group Publishing, to appear.

Baum-Snow, Nathaniel and Justin Marion. 2009. "The Effect of Low Income Housing Tax Credit Developments on Neighborhoods." *Journal of Public Economics* 93 (5-6): 654—666.

Carlson, Deven, Robert Haveman, Thomas Kaplan, and Barbara Wolfe. 2012. "Long-Term Effects of Public Low-Income Housing Vouchers on Neighborhood Quality and Household Composition." *Journal of Housing Economics* 21 (2): 101—120.

Chetty, Raj and Nathaniel Hendren. 2015. "The Impacts of Neighborhoods on Intergenerational Mobility: Childhood Exposure Effects and County-Level Estimates." Working Paper, Harvard University.

Collinson, Robert, Ingrid Gould Ellen, and Jens Ludwig. 2015. "Low-Income Housing Policy." NBER Working Paper No. 21071.

Congressional Budget Office. 2015. "Federal Housing Assistance for Low-Income Households." Congress of the United States Congressional Budget Office Publication 50782. https://www.cbo.gov/publication/50782.

Diamond, Rebecca and Tim McQuade. 2015. "Who Wants Affordable Housing in their Backyard? An Equilibrium Analysis of Low Income Property Development." Working Paper, Stanford University.

Eriksen, Michael D. 2009. "The Market Price of Low-Income Housing Tax Credits." *Journal of Urban Economics* 66 (2): 141—149.

Eriksen, Michael D. and Amanda Ross. 2015 "Housing Vouchers and the Price of Rental Housing." *American Economic Journal: Economic Policy* 7 (3): 154—176.

Freedman, Matthew and Tamara McGavock. 2015. "Low-Income Housing Development, Poverty Concentration, and Neighborhood Inequality." *Journal of Policy Analysis and Management*, Forthcoming.

Green, Richard K., Stephen Malpezzi, and Kiat-Ying Seah. 2002. "Low Income Housing Tax Credit Housing Developments and Property Values." The Center for Urban Land Economics Research, University of Wisconsin.

Hollar, Michael and Kurt Usowski. 2007. “Low-Income Housing Tax Credit Qualified Census Tracts.” *Cityscape* 9 (3): 152—159.

Horn, Keren M. and Katherine M. O’Regan. 2011. “The Low Income Housing Tax Credit and Racial Segregation.” *Housing Policy Debate* 21 (3): 443—473.

Horn, Keren M. and Katherine M. O’Regan. 2013. “What Can We Learn About the Low Income Housing Tax Credit Program by Looking at the Tenants.” *Housing Policy Debate* 23 (3): 597—613.

Jackson, Osborne and Laura Kiwano. 2015. “Do Increases in Subsidized Housing Reduce the Incidence of Homelessness? Evidence from the Low-Income Housing Tax Credit.” Unpublished Working Paper.

Jacob, Brian A. and Jens Ludwig 2012. “The Effects of Housing Assistance on Labor Supply: Evidence from a Voucher Lottery.” *American Economic Review* 102 (1): 272—304.

Jacob, Brian A., Jens Ludwig, and Douglas L. Miller. 2013. “The Effects of Housing and Neighborhood Conditions on Child Mortality.” *Journal of Health Economics* 32 (1): 195—206.

Jacob, Brian A., Max Kasputin, and Jens Ludwig. 2015. “The Impact of Housing Assistance on Child Outcomes: Evidence from a Randomized Housing Lottery.” *Quarterly Journal of Economics* 130 (1): 465—506.

Khadduri, Jill, Carissa Climaco, and Kimberly Burnett. 2012. “What Happens to Low-Income Housing Tax Credit Properties at Year 15 and Beyond?” U.S. Department of Housing and Urban Development, Office of Policy Development and Research.

Lee, Chang-Moo, Dennis P. Culhane, and Susan M. Wachter. 1999. "The Differential Impacts of Federally Assisted Housing Programs on Nearby Property Values: A Philadelphia Case Study." *Housing Policy Debate* 10 (1): 75—93.

Lee, David and Thomas Lemieux. 2010. “Regression Discontinuity Designs in Economics.” *Journal of Economic Literature* 48 (2): 281—355.

Olsen, Edgar O., Catherine A. Tyler, Jonathan W. King, and Paul E. Carrillo. 2005. “The Effects of Different Types of Housing Assistance on Earnings and Employment.” *Cityscape* 8 (2): 163—187.

Rohe, William and Lance Freeman. 2001. “Assisted Housing and Residential Segregation: The Role of Race and Ethnicity in the Siting of Assisted Housing Developments.” *Journal of the American Planning Association* 67 (3): 279—292.

Papay, John P., John B. Willett, and Richard J. Murnane. 2011. “Extending the Regression-Discontinuity Approach to Multiple Assignment Variables.” *Journal of Econometrics* 161 (2): 203—207.

Schwartz, Amy Ellen, Ingrid Gould Ellen, Ioan Voicu, and Michael H. Schill. 2006. “The External Effects of Place-Based Subsidized Housing.” *Regional Science & Urban Economics* 36 (6): 679–-707.

Sinai, Todd and Joel Waldfogel. 2005. “Do Low-Income Housing Subsidies Increase the Occupied Housing Stock?” *Journal of Public Economics* 89 (1-2): 2137—2164.

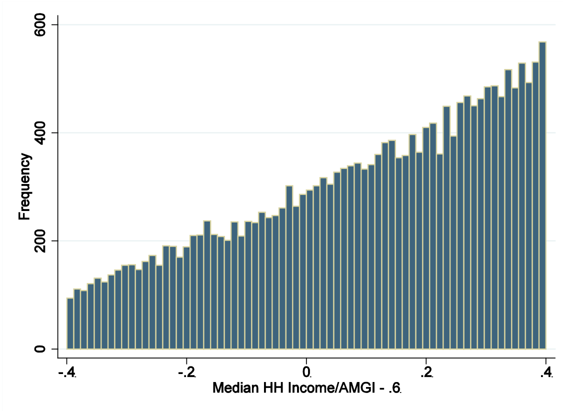
Susin, Scott. 2005. “Longitudinal Outcomes of Subsidized Housing Recipients in Matched Survey and Administrative Data.” *Cityscape* 8 (2): 189—218.

Wagner, Deborah and Mary Layne. 2014. “The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications' (CARRA) Record Linkage Software.” Center for Administrative Records Research and Applications Working Paper 2014-01, U.S. Census Bureau.

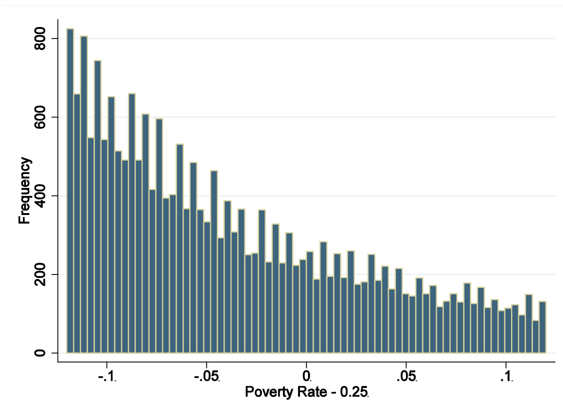
Woo, Ayoung, Kenneth Joh, and Shannon Van Zandt. 2014. “Impacts of the Low-Income Housing Tax Credit Program on Neighborhood Housing Turnover.” *Urban Affairs Review*: 1—33.

Figure 1: Marginal Distributions of Running Variables

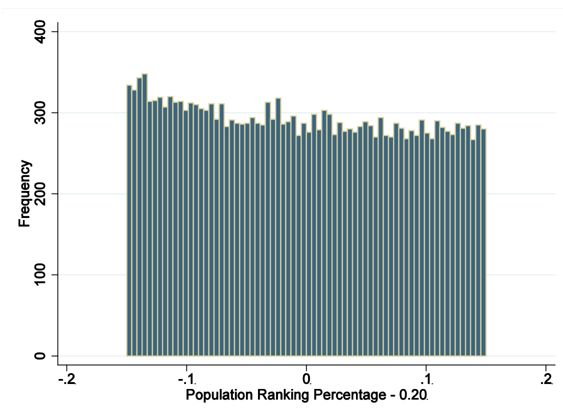
Panel A: Eligibility



Panel B: Poverty Rate



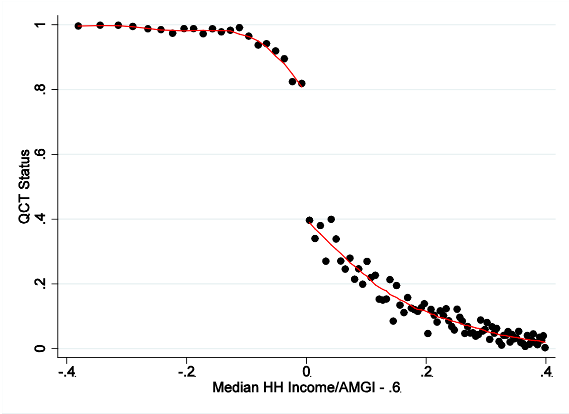
Panel C: Population Cap Ranking



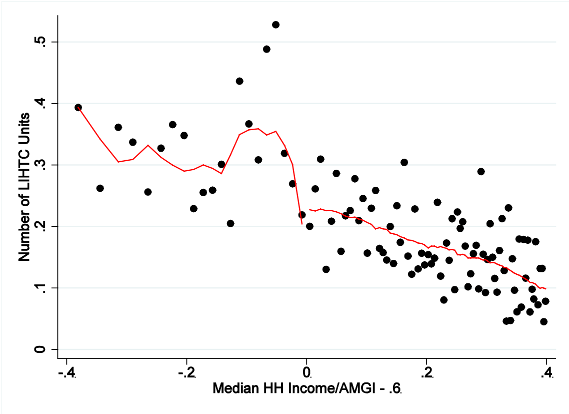
Source: Linked Census 2000 long form and 2010 Census short form microdata. Graphs plot histograms of a given running variable within the baseline bandwidths used to construct the estimates in Table 3.

Figure 2: LIHTC Development by Median Income of Census Tract Relative to 60% AMGI

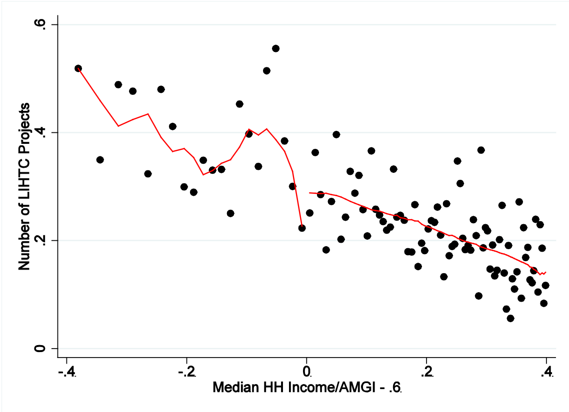
Panel A: QCT Status



Panel B: Number of LIHTC Units



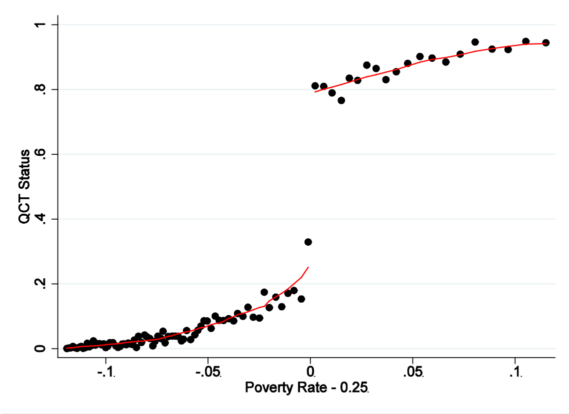
Panel C: Number of LIHTC Projects



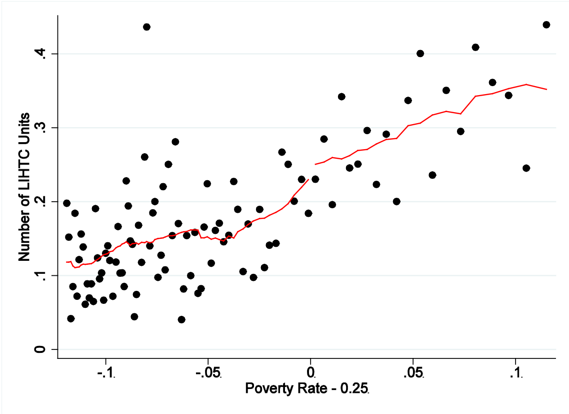
Source: Linked Census 2000 long form and 2010 Census short form microdata. Dots represent averages for 100 percentiles of the running variable within a given bandwidth. Red lines represent local linear regressions estimated separately on each side of the cutoff.

Figure 3: LIHTC Development by Poverty Rate of Census Tract

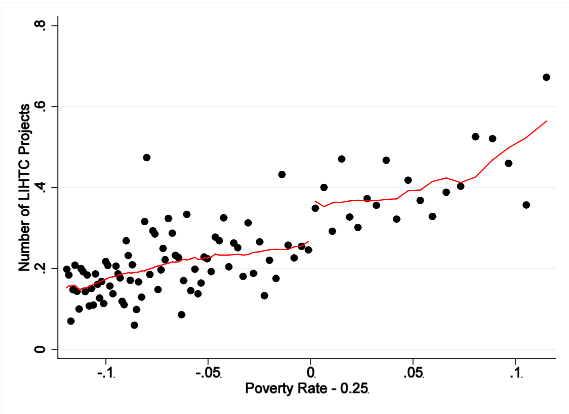
Panel A: QCT Status



Panel B: Number of LIHTC Units



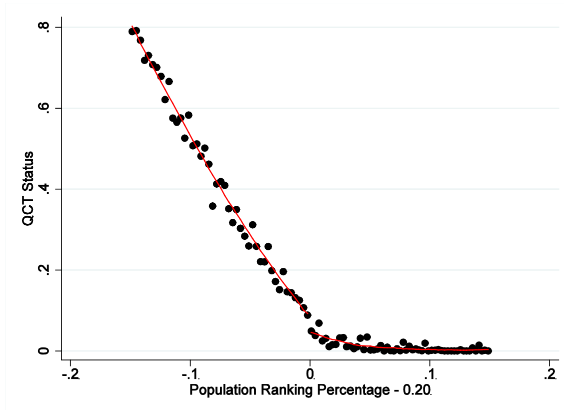
Panel C: Number of LIHTC Projects



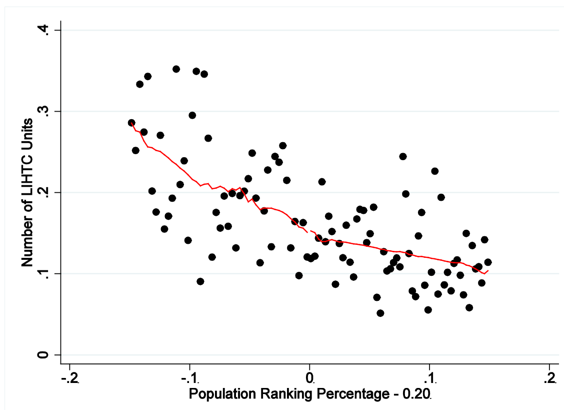
Source: Linked Census 2000 long form and 2010 Census short form microdata. Dots represent averages for 100 percentiles of the running variable within a given bandwidth. Red lines represent local linear regressions estimated separately on each side of the cutoff.

Figure 4: LIHTC Development in Relationship to Population Cap

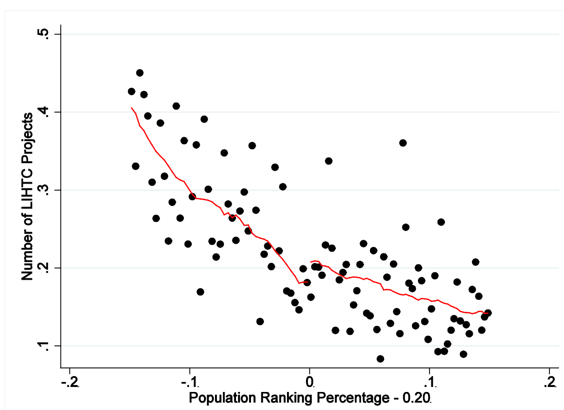
Panel A: QCT Status



Panel B: Number of LIHTC Units

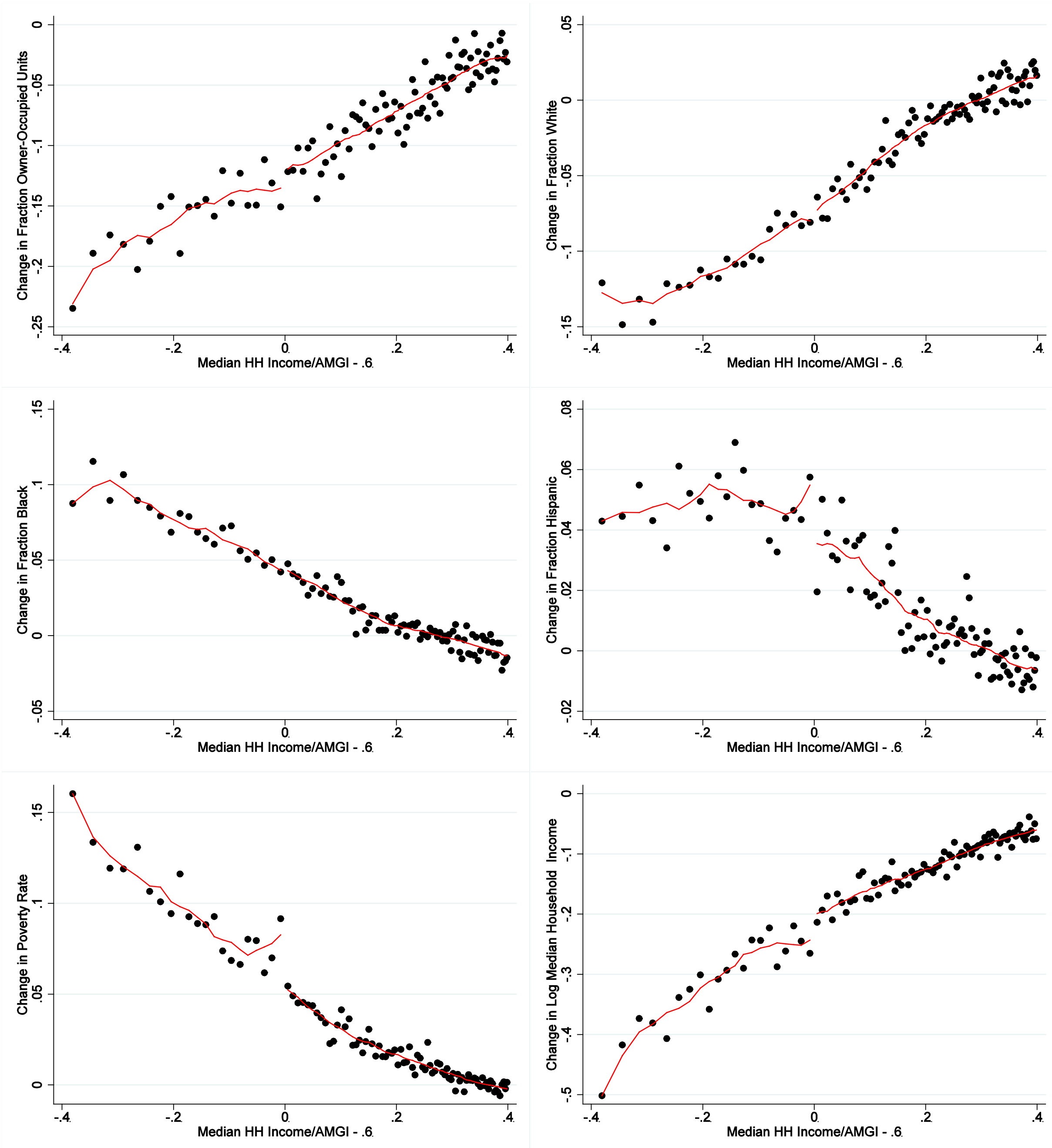


Panel C: Number of LIHTC Projects



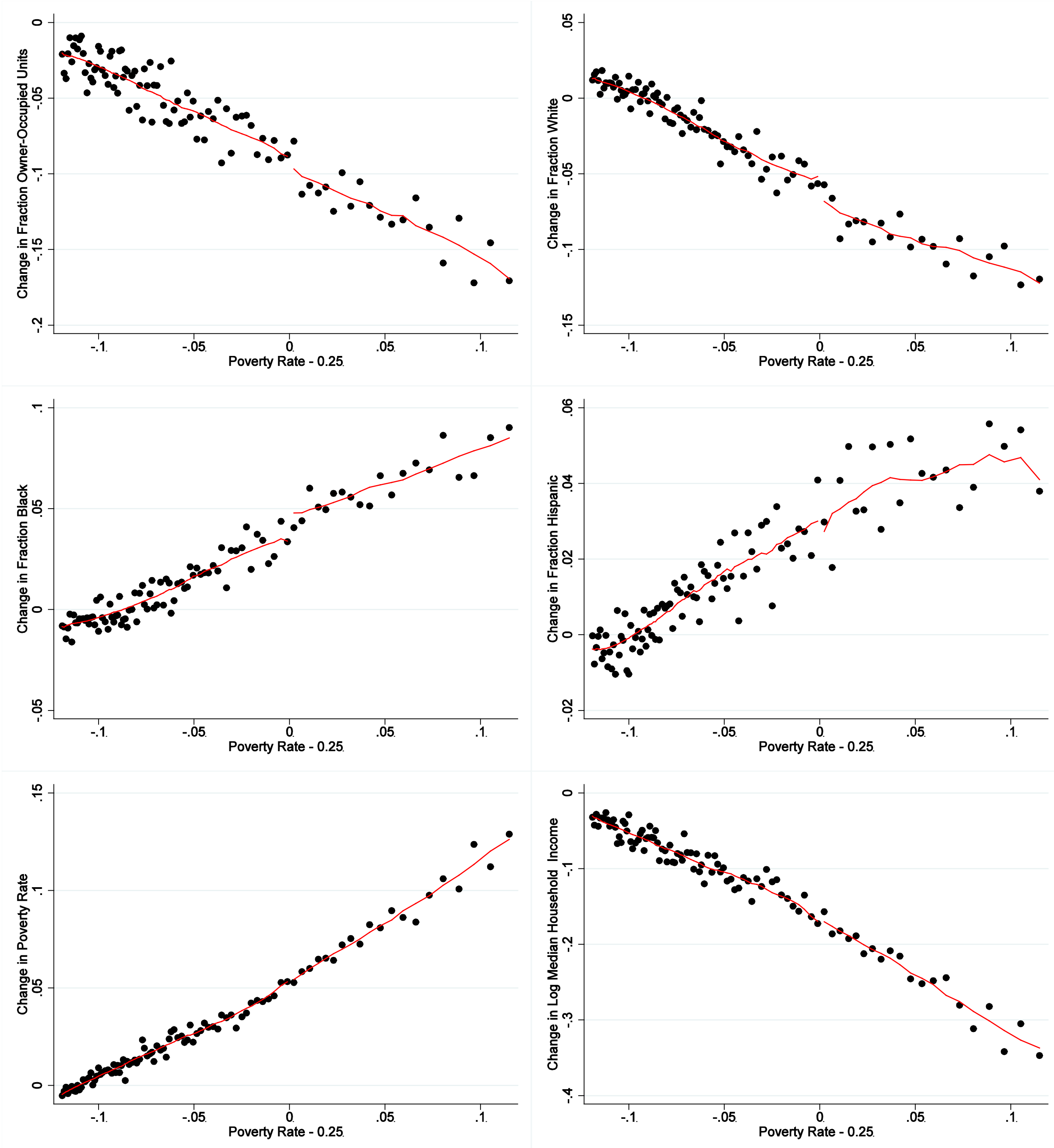
Source: Linked Census 2000 long form and 2010 Census short form microdata. Dots represent averages for 100 percentiles of the running variable within a given bandwidth. Red lines represent local linear regressions estimated separately on each side of the cutoff.

Figure 5: Change in Neighborhood Mobility Patterns by Eligibility Criterion



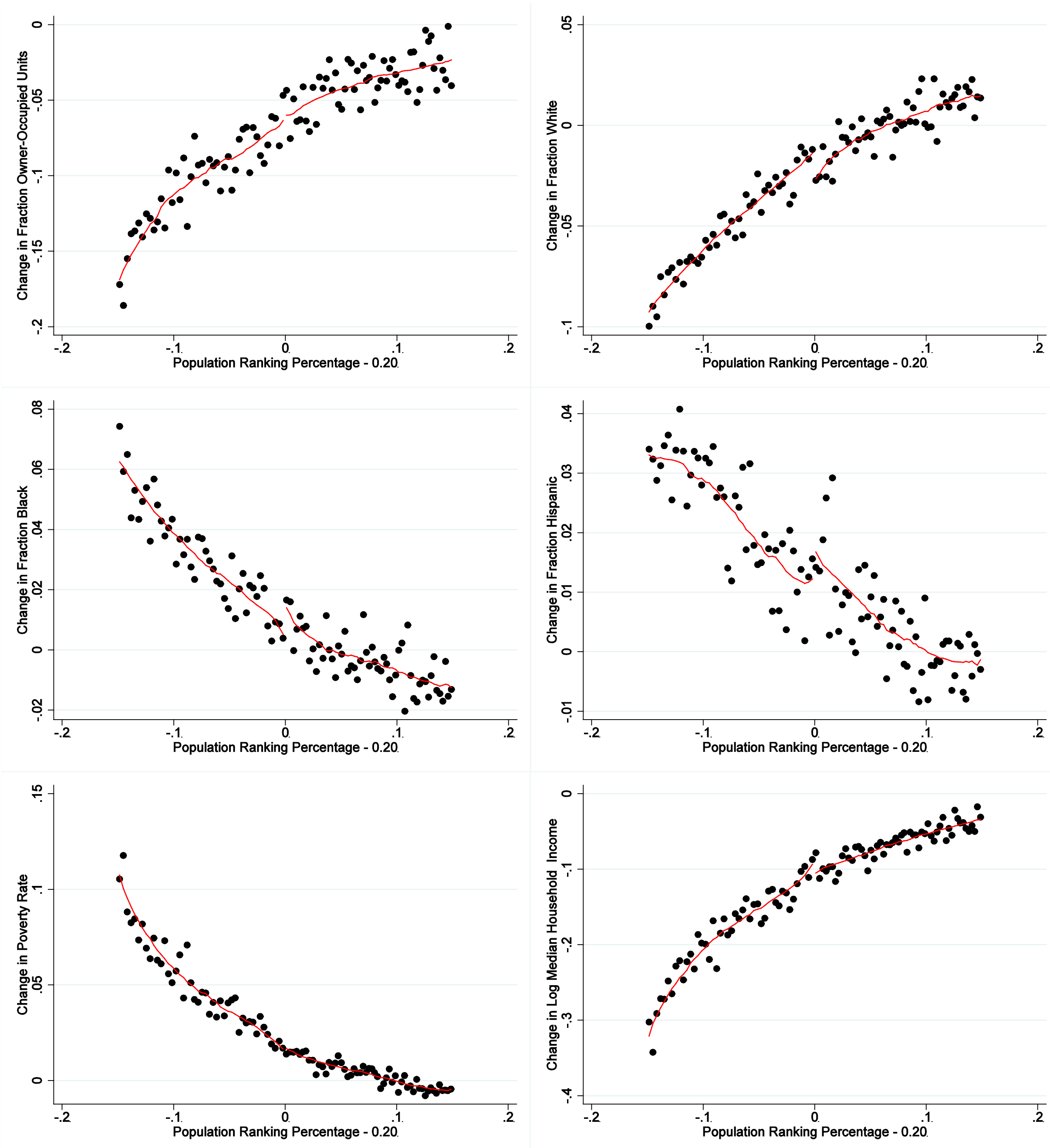
Source: Linked Census 2000 long form and 2010 Census short form microdata. Dots represent averages for 100 percentiles of the running variable. Red lines represent local linear regressions estimated separately on each side of the cutoff.

Figure 6: Change in Neighborhood Mobility Patterns by Poverty Rate



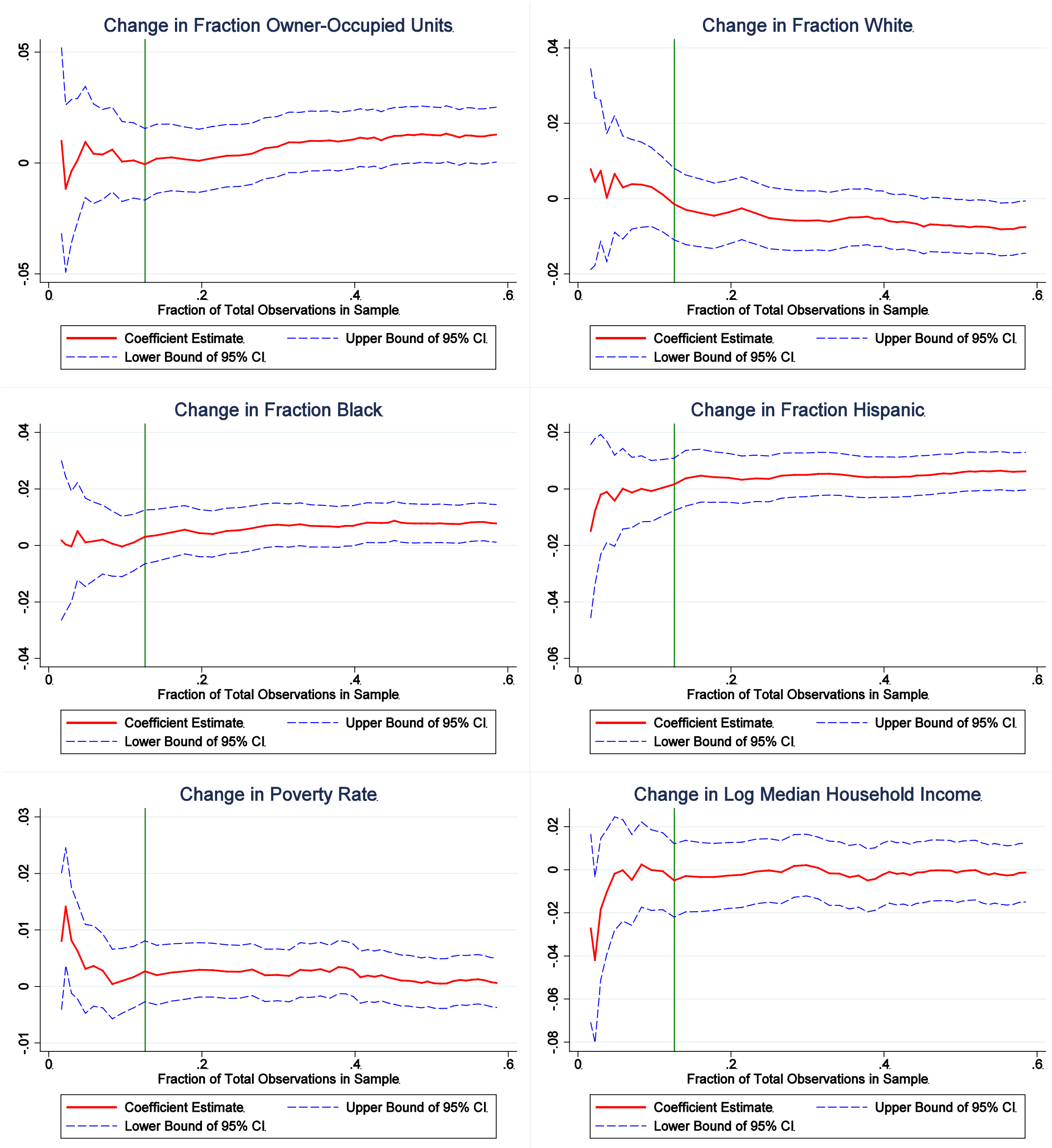
Source: Linked Census 2000 long form and 2010 Census short form microdata. Dots represent averages for 100 percentiles of the running variable. Red lines represent local linear regressions estimated separately on each side of the cutoff.

Figure 7: Change in Neighborhood Mobility Patterns by Population Cap Ranking



Source: Linked Census 2000 long form and 2010 Census short form microdata. Dots represent averages for 100 percentiles of the running variable. Red lines represent local linear regressions estimated separately on each side of the cutoff.

Figure 8: Sensitivity of RD Estimates to Bandwidth Choice



Source: Linked Census 2000 long form and 2010 Census short form microdata. All plots display estimated coefficients and corresponding confidence intervals for local linear regressions with a given outcome variable. Specifications are identical to those that produce the estimates in Table 4, with the exception that bandwidth is altered. Green vertical lines represent the bandwidth choice and corresponding estimates shown in Table 4.

Table 1: Descriptive Statistics

	Non-QCT	QCT
Panel A: Tract Characteristics:		
Fraction Black	0.074	0.326
Fraction White	0.829	0.469
Fraction Hispanic	0.088	0.235
Fraction <18 Years Old	0.255	0.275
Fraction >65 Years Old	0.128	0.108
HH Size	3.491	4.005
Log Median HH Income	11.036	10.307
Poverty Rate	0.091	0.332
Fraction Owner-Occupied Housing Units	0.668	0.350
Fraction Occupied Housing Units	0.838	0.789
Fraction of New Residents in 2010	0.613	0.542
Panel B: Changes in Base-Year Tract Characteristics for New Residents:		
Change in Fraction Own	0.032	-0.164
Change in Fraction Black	-0.019	0.071
Change in Fraction White	0.033	-0.101
Change in Fraction Hispanic	-0.016	0.036
Change in Poverty Rate	-0.021	0.112
Change in Log Median HH Income	0.071	-0.335
Panel C: Changes in Characteristics of New Tract 2000-2010:		
Change in Fraction Own 2000-2010	-0.041	-0.037
Change in Fraction Black 2000-2010	0.012	-0.019
Change in Fraction White 2000-2010	-0.025	0.041
Change in Fraction Hispanic 2000-2010	0.036	0.029
Change in Poverty Rate 2000-2010	0.037	0.027
Change in Log Median HH Income 2000-2010	-0.089	-0.072
N	17,262,285	1,513,410
N Clusters	54,353	9,262

Source: Linked Census 2000 long form and 2010 Census short form microdata, restricted to individuals who moved census tracts between 2000 and 2010. The unit of observation is the census tract. Cells include sample means.

Table 2: The Effect of QCT Status on LIHTC Development

	Dependent Variable		
	QCT Status	Number of LIHTC Units	Number of LIHTC Projects
Panel A: Polynomial Running Variable			
QCT Status	-	11.21 ***	0.1145***
	(-)	(3.90)	(0.0372)
N	-	2,884,623	2,884,623
N Clusters	-	11,043	11,043
Panel B: Eligibility Running Variable			
Estimated Discontinuity	0.4644***	10.25*	0.0710
	(0.0225)	(5.54)	(0.0560)
N	5,257,615	5,257,615	5,257,615
N Clusters	19,845	19,845	19,845
Panel C: Poverty Rate Running Variable			
Estimated Discontinuity	0.5764***	3.96	0.1078**
	(0.0218)	(4.73)	(0.0527)
N	6,050,620	6,050,620	6,050,620
N Clusters	20,327	20,327	20,327
Panel D: Population Cap Ranking Running Variable			
Estimated Discontinuity	0.0399***	1.57	-0.0140
	(0.0142)	(2.97)	(0.0346)
N	5,669,456	5,669,456	5,669,456
N Clusters	19,425	19,425	19,425

Source: Linked Census 2000 long form and 2010 Census short form microdata. The unit of observation is the person. Estimates are from local quadratic regressions using the following bandwidths: 0.4 for eligibility, 0.12 for poverty rate, and 0.15 for the cutoff ranking. Polynomial running variable regression uses the same bandwidths and includes a fully interacted quadratic specification of the three variables. Standard errors in parentheses are robust to clustering at the census tract level. *** indicates significance at the 0.01 level, ** at the 0.05 level, and * at the 0.10 level.

Table 3: Reduced Form Estimates of LIHTC Development on Neighborhood Mobility by Running Variable

	Dependent Variable					
	Change in Fraction Owner-Occupied	Change in Fraction Black	Change in Fraction White	Change in Fraction Hispanic	Change in Poverty Rate	Change in Log Median HH Income
Panel A: Eligibility Running Variable						
Estimated Discontinuity	0.0087	0.0007	-0.0054	-0.0007	-0.0216***	0.0404***
	(0.0117)	(0.0062)	(0.0064)	(0.0057)	(0.0057)	(0.0136)
N	4,255,255	4,255,255	4,255,255	4,255,255	4,255,255	4,255,255
N Clusters	19,845	19,845	19,845	19,845	19,845	19,845
Panel B: Poverty Rate Running Variable						
Estimated Discontinuity	-0.0104	0.0036	-0.0078	0.0009	0.0050*	-0.0125
	(0.0094)	(0.0052)	(0.0054)	(0.0049)	(0.0030)	(0.0107)
N	4,768,103	4,768,103	4,768,103	4,768,103	4,768,103	4,768,103
N Clusters	20,327	20,327	20,327	20,327	20,327	20,327
Panel C: Population Cap Ranking Running Variable						
Estimated Discontinuity	0.0092	0.0006	-0.0083*	0.0107***	-0.0067**	0.0073
	(0.0080)	(0.0038)	(0.0043)	(0.0041)	(0.0026)	(0.0079)
N	4,612,870	4,612,870	4,612,870	4,612,870	4,612,870	4,612,870
N Clusters	19,425	19,425	19,425	19,425	19,425	19,425

Source: Linked Census 2000 long form and 2010 Census short form microdata. The unit of observation is the person. Estimates are from local quadratic regressions using the following bandwidths: 0.4 for eligibility, 0.12 for poverty rate, and 0.15 for the cutoff ranking. Polynomial running variable regression uses the same bandwidths and includes a fully interacted quadratic specification of the three variables. Standard errors in parentheses are robust to clustering at the census tract level. *** indicates significance at the 0.01 level, ** at the 0.05 level, and * at the 0.10 level.

Table 4: Reduced Form Estimates of the Effect of QCT Status

	Dependent Variable						
	Change in Fraction Owner-Occupied	Change in Fraction Black	Change in Fraction White	Change in Fraction Hispanic	Change in Poverty Rate	Change in Log Median HH Income	Movers in Census Tract
Panel A: In Mobility							
QCT Status	-0.0006	0.0030	-0.0015	0.0017	0.0027	-0.0049	-
	(0.0082)	(0.0049)	(0.0049)	(0.0047)	(0.0027)	(0.0087)	-
N	2,319,192	2,319,192	2,319,192	2,319,192	2,319,192	2,319,192	-
N Clusters	11,043	11,043	11,043	11,043	11,043	11,043	-
Panel B: Neighborhood Characteristics							
QCT Status	-0.0044	-0.0061*	0.0099	0.0070*	-0.0044	-0.0021	-0.0050
	(0.0037)	(0.0033)	(0.0067)	(0.0039)	(0.0044)	(0.0115)	(0.0126)
N	2,319,192	2,319,192	2,319,192	2,319,192	2,319,192	2,319,192	2,319,192
N Clusters	11,043	11,043	11,043	11,043	11,043	11,043	11,043

Source: Linked Census 2000 long form and 2010 Census short form microdata. The unit of observation is the person. Estimates are from a linear regression including a fully interacted quadratic specification of the three running variables and using the following bandwidths: 0.4 for eligibility, 0.12 for poverty rate, and 0.15 for the cutoff ranking. Standard errors in parentheses are robust to clustering at the census tract level. *** indicates significance at the 0.01 level, ** at the 0.05 level, and * at the 0.10 level.

Table 5: IV Estimates of LIHTC Development

	Dependent Variable						
	Change in Fraction Owner-Occupied	Change in Fraction Black	Change in Fraction White	Change in Fraction Hispanic	Change in Poverty Rate	Change in Log Median HH Income	Movers in Census Tract
Panel A: In Mobility							
Number of LIHTC Projects	-0.0052	0.0266	-0.0128	0.0145	0.0235	-0.0431	-
	(0.0720)	(0.0431)	(0.0425)	(0.0414)	(0.0254)	(0.0774)	-
N	2,319,192	2,319,192	2,319,192	2,319,192	2,319,192	2,319,192	-
N Clusters	11,043	11,043	11,043	11,043	11,043	11,043	-
Panel B: Neighborhood Characteristics							
Number of LIHTC Projects	-0.0386	-0.0536	0.0862	0.0613	-0.0384	-0.0187	-0.0435
	(0.0338)	(0.0344)	(0.0630)	(0.0394)	(0.0401)	(0.1005)	(0.1125)
N	2,319,192	2,319,192	2,319,192	2,319,192	2,319,192	2,319,192	2,319,192
N Clusters	11,043	11,043	11,043	11,043	11,043	11,043	11,043

Source: Linked Census 2000 long form and 2010 Census short form microdata. The unit of observation is the person. Estimates are from a linear regression using QCT status as an instrument for a given endogenous variable, including a fully interacted quadratic specification of the three running variables, and using the following bandwidths: 0.4 for eligibility, 0.12 for poverty rate, and 0.15 for the cutoff ranking. Standard errors in parentheses are robust to clustering at the census tract level. *** indicates significance at the 0.01 level, ** at the 0.05 level, and * at the 0.10 level.

Table 6: Placebo Test using 2000 Tract Characteristics

	Dependent Variable					
	Fraction Owner-Occupied	Fraction Black	Fraction White	Fraction Hispanic	Poverty Rate	Log Median HH Income
Panel A: Reduced Form Estimates						
QCT Status	0.0200***	0.0048	0.0076	0.0042	-0.0000	-0.0001
	(0.0071)	(0.0124)	(0.0109)	(0.0108)	(0.0000)	(0.0073)
N	11,043	11,043	11,043	11,043	11,043	11,043
Panel B: IV Estimates						
Number of LIHTC Projects	0.3573*	0.0861	0.1358	0.0755	-0.0001	-0.0017
	(0.2080)	(0.2229)	(0.2052)	(0.1933)	(0.0002)	(0.1303)
N	11,043	11,043	11,043	11,043	11,043	11,043

Source: Linked Census 2000 long form and 2010 Census short form microdata aggregated to the tract level. Unit of observation is the census tract. Estimates are from a linear regression using QCT status as an instrument for a given endogenous variable, including a fully interacted quadratic specification of the three running variables, and using the following bandwidths: 0.4 for eligibility, 0.12 for poverty rate, and 0.15 for the cutoff ranking. Standard errors in parentheses are robust to heteroskedasticity. *** indicates significance at the 0.01 level, ** at the 0.05 level, and * at the 0.10 level.

Table 7: Heterogeneity in Results by Race

	Dependent Variable					
	Change in Fraction Owner-Occupied	Change in Fraction Black	Change in Fraction White	Change in Fraction Hispanic	Change in Poverty Rate	Change in Log Median HH Income
Panel A: White						
Number of LIHTC Projects	-0.0675	0.0618	-0.0626	0.0198	0.0447	-0.1082
	(0.0844)	(0.0486)	(0.0558)	(0.0445)	(0.0320)	(0.0932)
N	1,471,438	1,471,438	1,471,438	1,471,438	1,471,438	1,471,438
N Clusters	10,282	10,282	10,282	10,282	10,282	10,282
Panel B: African-American						
Number of LIHTC Projects	0.0140	0.0004	0.0244	-0.0247	-0.0024	-0.0258
	(0.0651)	(0.0833)	(0.0721)	(0.0361)	(0.0172)	(0.0600)
N	503,876	503,876	503,876	503,876	503,876	503,876
N Clusters	10,978	10,978	10,978	10,978	10,978	10,978

Source: Linked Census 2000 long form and 2010 Census short form microdata aggregated to the tract level. Estimates are from a linear regression using QCT status as an instrument for a given endogenous variable, including a fully interacted quadratic specification of the three running variables, and using the following bandwidths: 0.4 for eligibility, 0.12 for poverty rate, and 0.15 for the cutoff ranking. Standard errors in parentheses are robust to heteroskedasticity. *** indicates significance at the 0.01 level, ** at the 0.05 level, and * at the 0.10 level.

Table 8: Heterogeneity in Results by Household Income

	Dependent Variable					
	Change in Fraction Owner-Occupied	Change in Fraction Black	Change in Fraction White	Change in Fraction Hispanic	Change in Poverty Rate	Change in Log Median HH Income
Panel A: Quartile 1 (Lowest Income)						
Number of LIHTC Projects	0.0096	0.0034	0.0172	-0.0067	0.0100	-0.0074
	(0.0529)	(0.0413)	(0.0405)	(0.0345)	(0.0178)	(0.0587)
N	830,717	830,717	830,717	830,717	830,717	830,717
N Clusters	11,034	11,034	11,034	11,034	11,034	11,034
Panel B: Quartile 2						
Number of LIHTC Projects	-0.0195	0.0353	-0.0240	0.0133	0.0296	-0.0354
	(0.0675)	(0.0449)	(0.0444)	(0.0417)	(0.0249)	(0.0706)
N	605,136	605,136	605,136	605,136	605,136	605,136
N Clusters	11,026	11,026	11,026	11,026	11,026	11,026
Panel C: Quartile 3						
Number of LIHTC Projects	0.0189	0.0190	-0.0082	0.0327	0.0195	-0.0326
	(0.0886)	(0.0514)	(0.0513)	(0.0505)	(0.0293)	(0.0912)
N	485,682	485,682	485,682	485,682	485,682	485,682
N Clusters	11,017	11,017	11,017	11,017	11,017	11,017
Panel D: Quartile 4 (Highest Income)						
Number of LIHTC Projects	-0.0110	0.0641	-0.0457	0.0331	0.0330	-0.0734
	(0.1203)	(0.0645)	(0.0660)	(0.0681)	(0.0376)	(0.1403)
N	375,751	375,751	375,751	375,751	375,751	375,751
N Clusters	10,992	10,992	10,992	10,992	10,992	10,992

Source: Linked Census 2000 long form and 2010 Census short form microdata aggregated to the tract level, restricted to households with non-missing income information. Quartiles of income are based on total household income in the 2000 Census long form, and are based on all observations including those outside of the chosen bandwidth. Estimates are from a linear regression using QCT status as an instrument for a given endogenous variable, including a fully interacted quadratic specification of the three running variables, and using the following bandwidths: 0.4 for eligibility, 0.12 for poverty rate, and 0.15 for the cutoff ranking. Standard errors in parentheses are robust to heteroskedasticity. *** indicates significance at the 0.01 level, ** at the 0.05 level, and * at the 0.10 level.

Appendix A: Supplemental Results

Table A.1: Reduced Form Estimates of LIHTC Development on In Mobility by Running Variable – Linear Running Variable Specification

	Dependent Variable					
	Change in Fraction Owner-Occupied	Change in Fraction Black	Change in Fraction White	Change in Fraction Hispanic	Change in Poverty Rate	Change in Log Median HH Income
Panel A: Eligibility Running Variable						
Estimated Discontinuity	-0.0021 (0.0077)	0.0099 (0.0068)	0.0143*** (0.0042)	-0.0123*** (0.0038)	-0.0176*** (0.0037)	0.0135 (0.0091)
N	4,255,255	4,255,255	4,255,255	4,255,255	4,255,255	4,255,255
N Clusters	19,845	19,845	19,845	19,845	19,845	19,845
Panel B: Poverty Rate Running Variable						
Estimated Discontinuity	-0.0093 (0.0063)	0.0124** (0.0056)	-0.0132*** (0.0036)	0.0034 (0.0033)	0.0036* (0.0020)	-0.0083 (0.0072)
N	4,768,103	4,768,103	4,768,103	4,768,103	4,768,103	4,768,103
N Clusters	20,327	20,327	20,327	20,327	20,327	20,327
Panel C: Population Cap Ranking Running Variable						
Estimated Discontinuity	0.0035 (0.0052)	-0.0034 (0.0041)	-0.0061** (0.0029)	0.0042 (0.0026)	0.0025 (0.0018)	-0.0109** (0.0053)
N	4,612,870	4,612,870	4,612,870	4,612,870	4,612,870	4,612,870
N Clusters	19,425	19,425	19,425	19,425	19,425	19,425

Source: Linked Census 2000 long form and 2010 Census short form microdata. The unit of observation is the person. Estimates are from local linear regressions using the following bandwidths: 0.4 for eligibility, 0.12 for poverty rate, and 0.15 for the cutoff ranking. Polynomial running variable regression uses the same bandwidths and includes a fully interacted linear specification of the three variables. Standard errors in parentheses are robust to clustering at the census tract level. *** indicates significance at the 0.01 level, ** at the 0.05 level, and * at the 0.10 level.

Table A.2: Reduced Form Estimates of the Effect of QCT Status – Linear Running Variable

Dependent Variable

	Change in Fraction Owner-Occupied	Change in Fraction Black	Change in Fraction White	Change in Fraction Hispanic	Change in Poverty Rate	Change in Log Median HH Income	Movers in Census Tract
Panel A: In Mobility							
QCT Status	0.0166**	-0.0038	0.0069*	0.0037	-0.0026	0.0061	-
	(0.0072)	(0.0042)	(0.0042)	(0.0040)	(0.0022)	(0.0074)	-
N	2,319,192	2,319,192	2,319,192	2,319,192	2,319,192	2,319,192	-
N Clusters	11,043	11,043	11,043	11,043	11,043	11,043	-
Panel B: Neighborhood Characteristics							
QCT Status	-0.0050	-0.0056**	0.0147***	0.0029	-0.0056	-0.0004	-0.0084
	(0.0032)	(0.0028)	(0.0056)	(0.0034)	(0.0039)	(0.0100)	(0.0110)
N	2,319,192	2,319,192	2,319,192	2,319,192	2,319,192	2,319,192	2,319,192
N Clusters	11,043	11,043	11,043	11,043	11,043	11,043	11,043

Source: Linked Census 2000 long form and 2010 Census short form microdata. The unit of observation is the person. Estimates are from a linear regression including a fully interacted quadratic specification of the three running variables and using the following bandwidths: 0.4 for eligibility, 0.12 for poverty rate, and 0.15 for the cutoff ranking. Standard errors in parentheses are robust to clustering at the census tract level. *** indicates significance at the 0.01 level, ** at the 0.05 level, and * at the 0.10 level.

Table A.3: IV Estimates of LIHTC Development– Linear Running Variable

	Dependent Variable						
	Change in Fraction Owner-Occupied	Change in Fraction Black	Change in Fraction White	Change in Fraction Hispanic	Change in Poverty Rate	Change in Log Median HH Income	Movers in Census Tract
Panel A: In Mobility							
Number of LIHTC Projects	0.1696*	-0.0383	0.0700	0.0378	-0.0261	0.0625	-
	(0.0918)	(0.0444)	(0.0488)	(0.0417)	(0.0240)	(0.0786)	-
N	2,319,192	2,319,192	2,319,192	2,319,192	2,319,192	2,319,192	-
N Clusters	11,043	11,043	11,043	11,043	11,043	11,043	-
Panel B: Neighborhood Characteristics							
Number of LIHTC Projects	-0.0503	-0.0569	0.1486**	0.0298	-0.0571	-0.0043	-0.0850
	(0.0352)	(0.0350)	(0.0703)	(0.0365)	(0.0430)	(0.1010)	(0.1174)
N	2,319,192	2,319,192	2,319,192	2,319,192	2,319,192	2,319,192	2,319,192
N Clusters	11,043	11,043	11,043	11,043	11,043	11,043	11,043

Source: Linked Census 2000 long form and 2010 Census short form microdata. The unit of observation is the person. Estimates are from a linear regression using QCT status as an instrument for a given endogenous variable, including a fully interacted quadratic specification of the three running variables, and using the following bandwidths: 0.4 for eligibility, 0.12 for poverty rate, and 0.15 for the cutoff ranking. Standard errors in parentheses are robust to clustering at the census tract level. *** indicates significance at the 0.01 level, ** at the 0.05 level, and * at the 0.10 level.

Table A.4: Results Restricted to Non-DDA Areas

	Dependent Variable					
	Change in Fraction Owner-Occupied	Change in Fraction Black	Change in Fraction White	Change in Fraction Hispanic	Change in Poverty Rate	Change in Log Median HH Income
Panel A: Reduced Form Estimates						
QCT Status	0.0003	0.0030	-0.0030	0.0044	0.0019	0.0009
	(0.0090)	(0.0053)	(0.0053)	(0.0048)	(0.0030)	(0.0094)
N	1,956,947	1,956,947	1,956,947	1,956,947	1,956,947	1,956,947
N Clusters	9,142	9,142	9,142	9,142	9,142	9,142
Panel B: IV Estimates						
Number of	0.0022	0.0227	-0.0223	0.0329	0.0143	0.0065
LIHTC Projects	(0.0672)	(0.0403)	(0.0399)	(0.0370)	(0.0230)	(0.0699)
N	1,956,947	1,956,947	1,956,947	1,956,947	1,956,947	1,956,947
N Clusters	9,142	9,142	9,142	9,142	9,142	9,142

Source: Linked Census 2000 long form and 2010 Census short form microdata. The unit of observation is the person. Gentrification is defined as real median household income in the census tract increasing between 2000 and 2010. Estimates are from a linear regression using QCT status as an instrument for a given endogenous variable, including a fully interacted quadratic specification of the three running variables, and using the following bandwidths: 0.4 for eligibility, 0.12 for poverty rate, and 0.15 for the cutoff ranking. Standard errors in parentheses are robust to clustering at the census tract level. *** indicates significance at the 0.01 level, ** at the 0.05 level, and * at the 0.10 level.

Table A.5: Placebo Test using 2000 Tract Characteristics– Linear Running Variable

	Dependent Variable					
	Fraction Owner-Occupied	Fraction Black	Fraction White	Fraction Hispanic	Poverty Rate	Log Median HH Income
Panel A: Reduced Form Estimates						
QCT Status	0.0351***	0.0113	0.0073	0.0023	-0.0000	-0.0141**
	(0.0063)	(0.0108)	(0.0095)	(0.0091)	(0.0000)	(0.0061)
N	11,043	11,043	11,043	11,043	11,043	11,043
Panel B: IV Estimates						
Number of LIHTC Projects	0.5368***	0.1725	0.1120	0.0345	-0.0000	-0.2158*
	(0.1949)	(0.1726)	(0.1502)	(0.1390)	(0.0002)	(0.1137)
N	11,043	11,043	11,043	11,043	11,043	11,043

Source: Linked Census 2000 long form and 2010 Census short form microdata aggregated to the tract level. Unit of observation is the census tract. IV estimates are from a linear regression using QCT status as an instrument for a given endogenous variable, including a fully interacted linear specification of the three running variables, and using the following bandwidths: 0.4 for eligibility, 0.12 for poverty rate, and 0.15 for the cutoff ranking. Standard errors in parentheses are robust to heteroskedasticity. *** indicates significance at the 0.01 level, ** at the 0.05 level, and * at the 0.10 level.

Table A.6: Neighborhood Mobility Results by Age Group

	Dependent Variable					
	Change in Fraction Owner-Occupied	Change in Fraction Black	Change in Fraction White	Change in Fraction Hispanic	Change in Poverty Rate	Change in Log Median HH Income
Panel A: Age 0-22 in 2000						
Number of LIHTC Projects	-0.0275	0.0303	-0.0208	-0.0062	0.0440	-0.1051
	(0.1033)	(0.0524)	(0.0535)	(0.0496)	(0.0365)	(0.1110)
N	1,047,442	1,047,442	1,047,442	1,047,442	1,047,442	1,047,442
N Clusters	11,038	11,038	11,038	11,038	11,038	11,038
Panel B: Age 23-45 in 2000						
Number of LIHTC Projects	0.0078	0.0114	0.0115	0.0332	0.0241	-0.0314
	(0.0604)	(0.0484)	(0.0469)	(0.0485)	(0.0221)	(0.0674)
N	793,708	793,708	793,708	793,708	793,708	793,708
N Clusters	11,037	11,037	11,037	11,037	11,037	11,037
Panel C: Age 46+ in 2000						
Number of LIHTC Projects	0.0167	0.0282	-0.0126	0.0217	-0.0034	0.0338
	(0.0456)	(0.0302)	(0.0312)	(0.0300)	(0.0194)	(0.0646)
N	478,037	478,037	478,037	478,037	478,037	478,037
N Clusters	11,017	11,017	11,017	11,017	11,017	11,017

Source: Linked Census 2000 long form and 2010 Census short form microdata. The unit of observation is the person. Estimates are from a linear regression using QCT status as an instrument for a given endogenous variable, including a fully interacted linear specification of the three running variables, and using the following bandwidths: 0.4 for eligibility, 0.12 for poverty rate, and 0.15 for the cutoff ranking. Standard errors in parentheses are robust to clustering at the census tract level. *** indicates significance at the 0.01 level, ** at the 0.05 level, and * at the 0.10 level.

Table A.7: Heterogeneity by Homeownership

	Dependent Variable					
	Change in Fraction Owner-Occupied	Change in Fraction Black	Change in Fraction White	Change in Fraction Hispanic	Change in Poverty Rate	Change in Log Median HH Income
Panel A: Owners						
Number of LIHTC Projects	-0.0183	0.0411	-0.0360	0.0103	0.0198	-0.0310
	(0.0856)	(0.0404)	(0.0452)	(0.0372)	(0.0321)	(0.0944)
N	1,239,584	1,239,584	1,239,584	1,239,584	1,239,584	1,239,584
N Clusters	11,042	11,042	11,042	11,042	11,042	11,042
Panel B: Renters						
Number of LIHTC Projects	0.0340	0.0014	0.0309	0.0164	0.0217	-0.0406
	(0.0697)	(0.0575)	(0.0558)	(0.0531)	(0.0219)	(0.0760)
N	1,079,608	1,079,608	1,079,608	1,079,608	1,079,608	1,079,608
N Clusters	11,034	11,034	11,034	11,034	11,034	11,034

Source: Linked Census 2000 long form and 2010 Census short form microdata. The unit of observation is the person. Gentrification is defined as real median household income in the census tract increasing between 2000 and 2010. Estimates are from a linear regression using QCT status as an instrument for a given endogenous variable, including a fully interacted quadratic specification of the three running variables, and using the following bandwidths: 0.4 for eligibility, 0.12 for poverty rate, and 0.15 for the cutoff ranking. Standard errors in parentheses are robust to clustering at the census tract level. *** indicates significance at the 0.01 level, ** at the 0.05 level, and * at the 0.10 level.

Table A.8: LIHTC-Induced Mobility and Gentrification

	Dependent Variable					
	Change in Fraction Owner-Occupied	Change in Fraction Black	Change in Fraction White	Change in Fraction Hispanic	Change in Poverty Rate	Change in Log Median HH Income
Panel A: Gentrifying Neighborhoods						
Number of LIHTC Projects	-0.1030	-0.0477	0.0681	0.0056	0.0346	-0.0417
	(0.1642)	(0.0857)	(0.0992)	(0.1038)	(0.0507)	(0.1612)
N	651,487	651,487	651,487	651,487	651,487	651,487
N Clusters	3,162	3,162	3,162	3,162	3,162	3,162
Panel B: Not Gentrifying Neighborhoods						
Number of LIHTC Projects	0.0284	0.0655	-0.0510	0.0133	0.0209	-0.0397
	(0.0841)	(0.0562)	(0.0525)	(0.0421)	(0.0291)	(0.0860)
N	1,666,976	1,666,976	1,666,976	1,666,976	1,666,976	1,666,976
N Clusters	7,881	7,881	7,881	7,881	7,881	7,881

Source: Linked Census 2000 long form and 2010 Census short form microdata, restricted to tracts that can be linked across time. The unit of observation is the person. Gentrification is defined as real median household income in the census tract increasing between 2000 and 2010. Estimates are from a linear regression using QCT status as an instrument for a given endogenous variable, including a fully interacted linear specification of the three running variables, and using the following bandwidths: 0.4 for eligibility, 0.12 for poverty rate, and 0.15 for the cutoff ranking. Standard errors in parentheses are robust to clustering at the census tract level. *** indicates significance at the 0.01 level, ** at the 0.05 level, and * at the 0.10 level.

Appendix B: Data Appendix

B.1 Construction of Data Set

The long form of the 2000 decennial census forms the basis of the data set. These data contain detailed information for roughly 1 out of 6 households in the United States for the year 2000. These data are matched to the 2010 census using a probabilistic matching routine that takes into account birthdate, name, and geographic location (Wagner and Layne 2014). The discussion below contains a brief description of the matching process, as well as a description of how this process affects the demographic composition of the data set.

Table B.1 describes the creation of the person-level data set used in the analysis. The analysis data set is composed of all individuals who are able to be linked between the 2000 and 2010 censuses. The failure to link between the censuses is largely due to deaths and immigration between the censuses, but in part is due to error in the linkage process. For further information on how this affects the composition of the data set, refer to Rastogi and O'Hara (2012). After the linkage between the 2000 and 2010 decennial censuses, the data set is further restricted to individuals who can have their location linked to Census 2000 geographic definitions. This geographic conversion is performed using the Master Address File, an internal file at the Census Bureau that contains address information for all households in the Census. We then lose small numbers of individuals because their tract could not be linked to public Census 2000 data used to construct QCT classification information at either the tract or block group level.

All analyses in the paper use 2000 tract definitions. Note that for many of the mobility analyses, the sample is restricted to only those individuals who move census tracts between 2000 and 2010.

Table B.1: Summary of Sample Creation

Step	N	Percent Retained	Cumulative Percent Retained
Census 2000 Long Form	45,088,538		
Linked Data 2000-2010	32,052,398	71.09%	71.09%
Linked to Census 2000 Geographies	31,161,393	88.48%	69.11%
Linked to Tract-level QCT Data	30,238,508	97.04%	67.06%
Linked to Baseline Tract-Level Controls	30,200,699	99.87%	66.98%

Source: Linked Census 2000 long form and 2010 Census short form microdata. The unit of observation is the person. Note that number of observations differs from that in Table 1 of the main text, because Table 1 is restricted to only individuals who moved census tracts between 2000 and 2010.

B.2 Linkage Process

The matching process to link individuals between the 2000 and 2010 censuses first matches records to administrative records drawn from the Social Security Administration Numerical Identification File (Numident) and Internal Revenue Service Tax Records. Each person is assigned a unique protected identification key (PIK), which forms the link between decennial census records. The matching relies primarily on name, birthdate, and address data. Note that not all individuals in a given census are assigned a PIK. Rastogi and O'Hara (2012) and Layne et al. (2013) contain detailed descriptions of the match performance, but it is worth briefly noting a few points. First, because the procedure matches both files first to administrative records prior to merging the files together, the universe of individuals who theoretically could be linked encompasses all individuals with a Social Security Number as well as anyone who filed taxes in a given year.[36] In addition, zip code is a key matching field for each data set, allowing households to move between 2000 and 2010 without corrupting the linkage process.

[36] A small fraction of observations receive the same PIK after matching. We allow for this to occur, provided that we can observe individuals in both Censuses. Results are not sensitive to the use of de-duplicated data.

References

Rastogi, Sonya and Amy O'Hara. 2012. "2010 Census Match Study." Center for Administrative Records Research and Applications Report, U.S. Census Bureau.

Wagner, Deborah and Mary Layne. 2014. "The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications' (CARRA) Record Linkage Software." Center for Administrative Records Research and Applications Working Paper 2014-01, U.S. Census Bureau.