The Role of Census Tract Boundaries

Aaron Scholl*

Job Market Paper

March 31, 2021

Abstract

An important but less well-known feature of census tract geography is the role boundaries play in determining low-income, government-subsidized neighborhoods. Every ten years, in preparation for the Decennial Census, these boundaries are updated to accommodate population changes, and little is known to what extent these boundary changes affect the racial composition of neighborhoods. This paper investigates short- and long-run implications of neighborhoods that "split", exploiting Decennial Censuses from 1980 to 2010. Using difference-in-differences and event study designs, I find that "split" census tracts increase in their proportion of Black residents and these effects persist decades. Common trends analysis suggests these effects cannot be explained by observable differences prior to each Decennial Census. Instead, I find evidence that the Low-Income Housing Tax Credit program may play a role in concentrating residents in areas with greater census tract delineation. My results suggest that census tract delineation may play an important role in shaping neighborhood dynamics.

Keywords: census tracts, neighborhoods, urbanization, racial inequality, housing

*Department of Economics, University of Nebraska–Lincoln; e-mail: aaron.scholl@huskers.unl.edu. I thank seminar participants in the Labor Reading Group at the University of Nebraska for helpful comments and suggestions.

1 Introduction

In a majority of social science research, the census tract has long been the unit of analysis when studying "neighborhoods" because of their relatively stable geographic structure. For instance, economists often use census tracts to study neighborhood characteristics (Card et al., 2008; Farrell and Lee, 2011; Galiani et al., 2015; Bayer et al., 2016; Sharkey, 2016), create levels of racial separation (Cutler and Glaeser, 1997; Collins and Margo, 2000; Cutler et al., 2008; Ananat, 2011; Bayer and McMillan, 2012; Owens, 2016), or to measure the causal impacts of neighborhoods on various economic outcomes (Sharkey, 2016; Chetty et al., 2016; Chetty and Hendren, 2018; Chetty et al., 2020). While the most common usage of census tracts may be to distinguish geographically-stable neighborhoods, census tract boundaries also function through important public policy channels.

The largest and longest-running federal program to encourage the creation of affordable housing, the Low-Income Housing Tax Credit (LIHTC), uses census tract data in order to determine low-income, qualifying neighborhoods (Office of Policy Development and Research, 2013). Historically, census tract boundaries have also been used to establish election districts, precincts, and wards, which state and local governments create for administering elections (U.S. Census Bureau, 1994). Because of the potential implications census tract boundaries may have throughout cities, incentives to delineate strategically may arise. In preparation for the Decennial Census, these boundaries are updated to accommodate population changes, and little is known to what extent these boundary changes impact neighborhood racial dynamics and economic opportunity.

In this paper, I exploit variation in census tract boundaries using two distinct data sets to describe how neighborhood dynamics, in terms of racial composition, have changed over both short- and long-run time horizons as neighborhoods continue to urbanize. To evaluate short-run impacts on neighborhood composition, I use annual data within the context of a difference-in-differences (DID) empirical design surrounding the 2010 Decennial Census. Specifically, I examine the annual racial composition of census tracts that split relative to those that don't, before and after the 2010 Decennial Census holding census tract boundaries fixed to their 2000 Decennial Census "envelopes". Similarly, long-run estimations use an event study approach with Decennial Census data from 1980 to 2010 to evaluate how census tract envelope racial composition changes each decade upon further delineation. I conclude by examining a potential mechanism through which census tract delineation may impact the racial composition of the neighborhood – the Low-Income Housing Tax Credit.

Central to the analysis is the configuration and conditions under which census tract boundaries change. When originally introduced and widely implemented for the 1980 Decennial Census, a typical census tract would split prior to the Census if the population surpassed 8,000 residents, and merge if residential population fell below 1,200 (U.S. Census Bureau, 1994).[1] Original census tracts of the 1970 and 1980 Decennial Censuses were intended to be relatively permanent statistical subdivisions that delineated the entirety of the United States, while future delineations were expected to fall predominantly within initial boundaries (U.S. Census Bureau, 2008).[2] These delineations often reflect geographical boundaries, and were historically used to aid census enumerators in visually depicting boundaries which enclosed their respective areas of data collection.[3]

The key identifying assumption in estimating causal impacts of census tract delineation is that trends in neighborhood racial composition would have evolved in the same way in neighborhoods that received additional census tract splits as they would at non-split neighborhoods within the same city, absent census tract boundary updating. Within this context, I provide evidence that there are no observable racial differences between neighborhoods that split and those that don't prior to each Decennial Census. While my results lack evidence of differential neighborhood racial compositions prior to boundary updates, or parallel pre-trends, like all DID designs, I cannot directly tests for differences in pre-trends in unobserved determinants of neighborhood racial composition.

The first key finding of the analysis is that relative to census tract envelopes that don't receive further delineation, "split" census tract envelopes have higher proportions of black residents after the resulting Decennial Census boundary updates, and these effects persist decades. In terms of short-run effects, I find that 2000 Decennial Census census tract envelopes that split in the 2010 Census increased in their proportion of Black residents by 0.26 percentage points relative to census tracts that did not split. At the same time, I find evidence that census tract envelopes that were split reduced the proportion of white residents in the neighborhood by 0.16 percentage points immediately following the 2010 Decennial Census. Long-run estimates suggest these effects amplify over time: twenty years beyond further census tract delineation results in the neighborhood proportion of Black residents increasing by 5.03 percentage points, while the proportion of white residents decreased 1.65 percentage points.

In exploring the LIHTC Qualified Census Tract mechanism, I find evidence of a concentration effect. My results suggest that low-income, qualifying census tracts are 1.6 percentage points more likely to receive multiple LIHTC-funded housing projects if the census tract envelope received further delineation in the 2010 Decennial Census, relative to similar low-income, qualifying census tracts that did not split. Investigating heterogeneous impacts suggests that the probability of multiple LIHTC-funded projects within a neighborhood increases by 0.6 percentage points for each additional split within the census tract envelope. Lastly, I

[1]These thresholds have not been updated since their introduction and continue to serve as a general guideline.

[2]This was intended to simplify and enhance the tracking process of data at the census tract-level over time. Henceforth, I refer to census tracts originating in 1970 and 1980 as the census tract envelope.

[3]There are cases of "special" census tracts where these guidelines are abandoned and populations are much larger. For example, some census tracts that contain large college campuses can be considered a special census tract due to the large populations within the neighborhood.

find evidence that splits within low-income, qualifying neighborhoods temporarily increases housing values while at the same time reduces neighborhood rental rates – features that are likely explained by improvements to the neighborhood housing stock and the designation of LIHTC income-restricted rental units, respectively.

The most innovative feature of this paper is the application of census tract boundaries beyond their use of geographically stable neighborhoods. Rather than only using census tracts as the neighborhood unit of analysis (e.g., Card et al. (2008); Chetty and Hendren (2018); Chetty et al. (2020)), my empirical approach moves us forward in evaluating how further delineation of these boundaries impacts outcomes of interest. The program developed by Logan et al. (2014) allows me to combine and distribute census tract boundaries and characteristics to various points in time. Thus, I am able to create a consistent shell, or "envelope", for studying the economics of census tracts.

More broadly this paper contributes to a relatively new strand of literature studying the economics of maps. Nagaraj and Stern (2020) provide a detailed discussion of the significance of map making in directing and developing economic activity dating back to the Middle Ages. Jubara et al. (2021) work with not-yet-released National Historical Geographic Information System (NHGIS) census block data dating back to 1980 boundaries to explore the history of racial segregation and integration of the Black population in metropolitan Minnesota. They find that overlaying historical racial covenants and Home Owners' Loan Corporation (HOLC) zones with census block data continues to leave persistently low or high proportions of Black residents in targeted neighborhoods. In a forthcoming paper, Aaronson et al. (Forthcoming) study the effects of the 1930s HOLC "redlining" maps on the long-run trajectories of urban neighborhoods. The authors find that the HOLC maps of the early 20th century had meaningful and lasting effects on the racial development of urban neighborhoods through housing availability and neighborhood disinvestment.

More narrowly defined this paper contributes to understanding the link between census tract configuration and affordable housing options. To the extent that public policy, such as the LIHTC Qualified Census Tracts program, rely on the delineation of census tracts for determining qualified neighborhoods, several researchers discuss that little to no work exists in evaluating the consequences of such policy (e.g., Hollar and Usowski (2007); Dawkins (2013); Tax Policy Center (2017); Scally et al. (2018); Reid (2019)). This work serves to lay the foundation in establishing the role census tracts play in determining neighborhood racial composition through the LIHTC. I study whether further census tract boundary delineation influences the probability that low-income, qualifying neighborhoods receive LIHTC funds, as well as other housing market characteristics also likely to be influenced by the development of LIHTC housing.

The rest of this paper is organized as follows: Section 2 provides background into census tract formation, who determines boundaries, and public policy channels; Section 3 describes the data sources and their implementations; Section 4 presents the empirical analysis of studying census tract delineation; Section 5 investigates whether the LIHTC program is related to census tract delineation; Section 6 concludes.

2 History and Implementation of the Census Tract

In this section I develop the history and intuition for evaluating census tract formation and boundary updates. I begin by formally discussing the historical timeline of census tracts in the U.S. Next, I describe how census tracts evolve in preparation for each Decennial Census, and provide a visual depiction of census tract boundary updates. I conclude with a discussion of who maintains and determines boundary delineation and potential channels through which the delineation process may create separation.

2.1 History of the Census Tract

As written in the U.S. Const. art. I, § 2.:

> *"Representatives and direct Taxes shall be apportioned among the several States which may be included within this Union, according to their respective Numbers [...] The actual Enumeration shall be made within three Years after the first Meeting of the Congress of the United States, and within every subsequent Term of ten Years, in such Manner as they shall by Law direct."*

Accordingly, the first census was carried out by U.S. marshals in 1790, consisted of six questions, and collected data on approximately 3.9 million people (U.S. Census Bureau, 2002). One hundred years later, Dr. Walter Laidlaw, the Executive Secretary of the New York City Census Committee, proposed that in order to study neighborhoods meaningfully it was necessary to have population data smaller than the borough or ward (U.S. Census Bureau, 2013). He argued these data collection areas should remain unchanged from census to census, and first referred to the established areas in New York City as "sanitary districts" in 1906; however beyond this, little progress would be made on the development of district-level data over the following 20 years.[4]

In preparation for the 1940 Decennial Census, several other large cities adopted the census tract as an official geographic unit for which data would be published. Census tracts covered large urban environments and block number areas (BNAs) covered several other suburban cities. In 1970 the number of BNAs increased and criteria of BNAs were updated to match those of the census tract (U.S. Census Bureau, 2013). By the 1990 Decennial Census, the entirety of the U.S. was delineated in census tract and block number areas, and

[4]Decennial Census committee members were skeptical of the use of the more detailed level data relative to the level of preparation necessary for delineation and collection of smaller geographic districts. For instance, in the 1930 Decennial Census, some cities developed additional committees to aid in processing census tract-level data, but were unable to raise the necessary funds for publication of the final tables.

in 2000 the BNA concept was retired in exchange for consistent census tract terminology. 2010 marked the 100th anniversary of the delineation of the first census tract boundaries in the United States.

2.2 Definition of the Census Tract

Original census tracts of the 1970 and 1980 Decennial Censuses were intended to be relatively permanent statistical subdivisions that delineated the entirety of the United States (U.S. Census Bureau, 1994). As neighborhoods urbanized, and populations grew, boundaries needed to adjust for accurate, timely, and sufficiently small geographic detail in the data collection process by census enumerators. Rather than newly configuring census tracts every Decennial Census, they instead split from the pre-existing census tract envelopes of earlier Decennial Censuses (U.S. Census Bureau, 2008). While small adjustments may be made to already existing boundaries, this feature enhances the tracking of census tract geography over time.[5] Standard census tracts are usually split beyond a residential population of 8,000 residents, and merged if resident population falls below 1,200.[6] Figure 1 illustrates the evolution of census tract boundaries since their featured adoption for the 1970 Decennial Census. These population thresholds have not been adjusted since their widespread implementation of the 1970 and 1980 Decennial Censuses. So, while neighborhoods have become more crowded over the last 50 years, boundary delineations are usually accepted as long as they fall within the specified range.

2.3 Who Determines the Boundaries of Census Tracts?

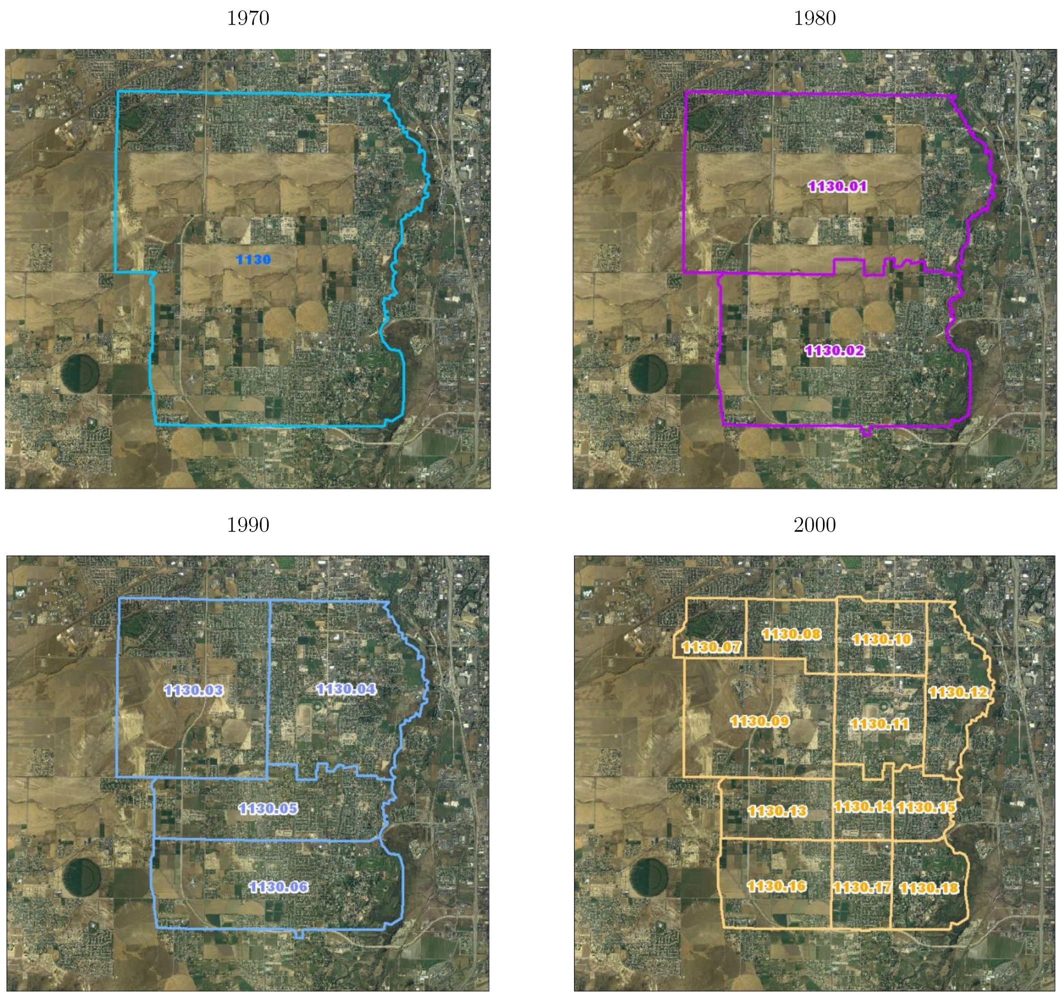
The Participant Statistical Areas Program (PSAP) through the U.S. Department of Commerce allows eligible participants to review, update, and delineate new census tracts prior to the upcoming Decennial Census. Participants may consist of any interested regional or local government entities, private organizations, and individuals familiar with the area. U.S. Census Bureau (2008) encourages participation for three primary reasons: (1) familiarity of local population changes and settlement patterns, (2) familiarity with unincorporated and emerging communities, and (3) resulting data to meet the needs of their communities. Because of the familiarity local participants have in determining census delineated places, the Census Bureau is able to meet many of the statistical and spatial needs required of the agency. U.S. Census Bureau (2008) discusses how data tabulated to these boundaries are used by local, state, and federal agencies and organizations for planning and funding purposes, as well as the private sector, academia, and the public.

Primary participants begin by agreeing to work with all interested parties so that the resulting plan

[5]These updates and small boundary corrections are allowed during intercensal population estimates.

[6]Special census tracts may be created for large special land use areas without housing or population (e.g., large public parks, forests, etc.) (U.S. Census Bureau, 2008)

Figure 1: Census Tract Boundary Changes



Source: U.S. Census Bureau (2013) This figure shows the evolution of census tract envelope 1130 of Salt Lake County, Utah, and illustrates how census tracts are split over time as the census tract becomes increasingly populated.

accommodates the needs and interests of governments, organizations, and individuals in the area.[7] Participants choose which counties they are interested in covering, and agree to review and update all Participant Statistical Areas within the specified county(ies). A final plan for each county is submitted to the Census Bureau where acceptable population criteria are verified[8]. Beyond approval from the Census Bureau, delineation files are inserted into the Topologically Integrated Geographic Encoding and Referencing (TIGER) Database.

2.4 Census Tracts in Public Policy

While U.S. Census Bureau (2008) describes that state and local governments may also use census tract geography in determining funding, the largest federal program to do so is the Low-Income Housing Tax Credit (LIHTC) Qualified Census Tracts program. Hollar and Usowski (2007) document the legislative history and methodology for designating and awarding LIHTC funds in qualified neighborhoods. The Tax Relief Act of 1986 created the LIHTC program, and the Omnibus Budget Reconciliation Act of 1989 amended the program to include additional incentives for rehabilitation and replacement of substandard rental housing in low-income areas, known as Qualified Census Tracts (QCTs).

Standard QCTs consist of tracts that have 50 percent of households with incomes below 60 percent of the Area Median Gross Income (AMGI), or have a poverty rate of 25 percent or more (Office of Policy Development and Research, 2013).[9] LIHTC funds for these neighborhoods are awarded on a state population basis from the Department of the Treasury, and distributed to housing developers through state housing finance agencies. There is no federal oversight beyond initial distribution, and housing agencies have a wide discretion in determining projects to award credits. Tax credits that are not purposed within the state are returned and redistributed for use in other areas. While changes in census tract boundaries should have no effect on the allocation of tax credits each state receives, boundary manipulation could enhance the number of LIHTC-funded affordable housing units available within low-income or high-poverty areas, or both.

Census tracts are first ranked in terms of lowest-income and highest poverty neighborhoods and second by their combined rank. The combined ranking determines the level of priority in awarding LIHTC funds. Thus, QCT envelopes potentially receive excessive funds through additional boundary delineation while concentrating affordable housing in the lowest-income or highest poverty areas. Several researchers suggest the use of census tracts in determining who receives LIHTC funds provides incentives for concentrating affordable housing in low-income or high-poverty areas (eg., Dawkins (2013); Tax Policy Center (2017);

[7] U.S. Census Bureau (2008) strongly encourages soliciting input from non-governmental organizations, academics, and interested individuals.

[8] Census tracts that fall outside of standard census tract guidelines must have appropriate documentation for the decision.

[9] Ellen and Horn (2018) discuss varying features of individual state Qualified Allocation Plans, in addition to QCTs, that further encourage the development and rehabilitation of neighborhoods that meet specified criteria.

Fischer (2018); Scally et al. (2018)). In fact Kennedy (2015), found that Texas was in violation of locating LIHTC-funded affordable housing units in neighborhoods of predominantly Black inner-city areas, and too few in predominantly white suburban neighborhoods.

While this work is primarily interested in how census tracts function within neighborhoods, historically census tract boundaries have also been used in determining voting districts. The boundaries of election districts after the 1970 Decennial Census were often drawn independent of Census Bureau geography, and relied on Census population counts. Because of the lengthy, and sometimes impossible, process in matching the features of state-provided election maps to features of Census Bureau geography, timely population counts were absent in updating districts (U.S. Census Bureau, 1994). Thus, the Census Bureau created the Election Precinct Program for the 1980 Decennial Census whereby state officials were instructed to delineate voting districts to coincide with sufficiently large Census Bureau geography – census tracts.[10] This ensured that the Census Bureau would be able to deliver accurate and timely population data by the end of the first quarter of the following year. As a result, incentives arise to potentially delineate boundaries strategically such that the resulting districts reflect the goals of a legislator, rather than a representative community.

3 Data Background and Usage

The data sets used in the analysis include census tract data from the Longitudinal Tract Data Base (LTDB) and National Historical Geographic Information System (NHGIS), as well as Low-Income Housing Tax Credit census tract data from the Department of Housing and Urban Development (HUD). In this section I describe each data set and individual sampling criteria in further detail and place them into context of the analysis.

3.1 Longitudinal Tract Data Base

Long-run census tract-level data come from the Longitudinal Tract Data Base (LTDB) (Logan et al., 2014), and allow researchers to construct a variety of data that stem from full count and sample data in the 1980-2000 Decennial Censuses and the 2008-2012 ACS 5-year data file. These census tract-level data include variables on population, race, income, education, and workforce characteristics. The key advantage in using the LTDB lies in its choice of census tract boundaries. Data sets are available to reflect either longitudinally consistent boundaries fixed to a single point in time, or the use of temporally-dependent census tract boundaries. Additionally, because the data are provided as raw population counts I am able to construct the characteristic

[10] At the end of the process, the Census Bureau returned all State-submitted maps to the States for their use in the redistricting process. Because the Census Bureau produced the election precinct tabulations as a special computer subfile, the data were not in any published report. Also of note, the Census Bureau kept no copies of these maps and did not show the boundaries of precincts on any 1980 census maps available to the public.(U.S. Census Bureau, 1994)

of interest through aggregation of count data, rather than weighting proportions to fit within larger levels of aggregation.[11] Using the Historical Delineation Files available from U.S. Census Bureau (2011), I aggregate the detailed raw census tract data up to the MSA in order to construct city-level estimates of population counts.[12] I restrict my analysis to urban MSAs defined as having a resident population of at least 100,000 individuals. This results in 303 unique MSAs in 2010, the most recent period used in the analysis.

I also make use of another feature of the LTDB when implementing the main DID empirical strategy. Logan et al. (2014) construct a program through the LTDB that matches 2010 Decennial Census census tract boundaries within their 2000 Decennial Census envelopes. This allows me to examine the racial composition of 2000 Decennial Census envelopes that split relative to those that don't, before and after the 2010 Decennial Census. The program contains information on how many splits occurred within the 2000 census tract boundaries and appropriately weights mean, median, and proportion variables by their base.[13]

3.2 National Historical Geographic Information System (NHGIS)

Additional short-run census tract-level population data come from the NHGIS available through IPUMS (Manson et al., 2019). The NHGIS provides access to summary statistics and Geographic Information System (GIS) files for the Decennial Censuses and other nationwide surveys at varying levels of geography. I make use of the ACS 5-year data files from 2005 to 2018 aggregated to the census tract level in my DID empirical analysis of census tract splits surrounding the 2010 Decennial Census. Each 5-year file provides estimates of the average characteristics over the relevant period, and I use these characteristics to reflect the midpoint in each ACS 5-year sample. For instance, the 2005-2009 ACS 5-year file characteristics reflect the midpoint, 2007, in my analysis. Thus, my DID analysis uses data from midpoints 2007 to 2016 reflecting ACS 5-year data files from 2005-2009 to 2014-2018. Variables of interest include the total population, white alone, and Black alone populations, as well as median rent and housing prices.

I apply the LTDB conversion program discussed above to the NHGIS population data in order to convert all data back to their 2000 Decennial Census envelopes, and calculate proportions white, Black, median rent, and median housing value using 2000 census tract envelope boundaries. Thus, I have data centered at the midpoints of ACS 5-year data files from 2007 to 2016, aggregated to match exact 2000 Decennial Census boundaries. The resulting sample consists of 456,926 envelope-year observations within the MSAs of interest.

[11] For example, I am able to construct census tract-level proportions using the raw census tract data. I can also aggregate the raw data to higher levels of geography, such as the MSA-level, without having to weight census tract proportions by their population, for instance.

[12] Census tracts are designed to fall entirely within counties, and counties fall entirely within MSA boundaries. Thus, census tracts aggregate nicely within MSAs.

[13] For example, the base variable for median household income would be the total number of households.

3.3 Department of Housing and Urban Development (HUD)

HUD's LIHTC database contains information on over 48,000 projects placed in service between 1987 and 2019. Variables used in the analysis include whether there exists multiple LIHTC-funded projects, the number of low-income restricted housing units, and whether units are new construction or rehabilitated. Data are available at the project level and aggregated to current census tract boundaries. I further apply the boundary conversion program developed by Logan et al. (2014) to adjust these characteristics to 2000 Decennial Census Boundaries. These data are merged with NHGIS data from 2007-2016 to analyze the impact of census tract delineation on the LIHTC characteristics of interest.

4 Empirical Analysis: the Effect of Census Tract Delineation

I begin my empirical analysis by studying the impact that additional boundaries have on the racial composition of census tract envelopes. The first section consists of a dynamic difference-in-differences (DID) empirical design estimating short-run impacts surrounding the 2010 Decennial Census. Then I present the main DID results empirically establishing significant effects *after* the 2010 Census, as well as heterogeneous impacts. I conclude by corroborating the results found in my short-run analysis with an alternate data set investigating long-run impacts.

4.1 Dynamic Difference-in-Differences

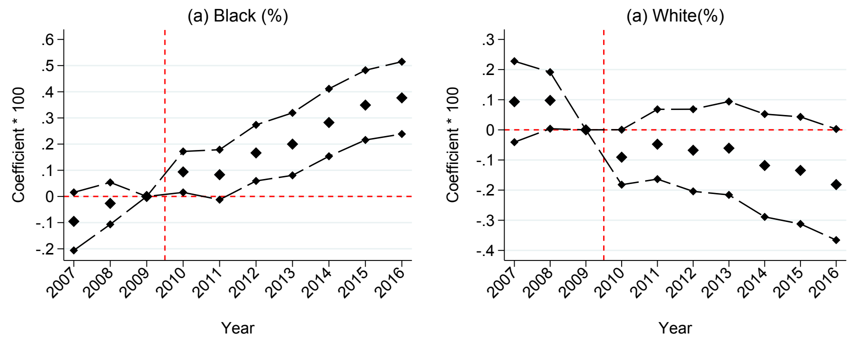
In order to characterize the impact that census tract delineation has on the racial composition of neighborhoods, I present a dynamic difference-in-differences analysis using the 2010 Decennial Census. Specifically, I examine the racial composition of 2000 Decennial Census census tract envelopes that split relative to those that don't, before and after the 2010 Decennial Census, from 2007 to 2016. My estimating equation of interest is:

$$y_{ict} = \alpha_0 + \delta treatment_i + \sum_{t=2007}^{2016} \gamma_t + \sum_{t=2007}^{2016} \delta_t treatment_i + x'_{it}\theta + \epsilon_{ict} \quad (1)$$

where the dependent variable y_{ict} is either the percent white or Black in census tract envelope i in city c at time t. The independent variable *treatment* equals 1 if the census tract envelope receives further delineation in the 2010 Decennial Census. The base year of analysis is 2009, the year before census tracts split. Additional controls in x_{it} include a quadratic function in population and city fixed effects. Standard errors are robust and clustered at the 2000 census tract envelope level.

The coefficient of interest in Equation 1 is δ_t and indicates the percentage point effect on the neighborhood racial composition in year t. The key assumption for identification is that trends in neighborhood racial

Figure 2: Effect of Census Tract Splits on Neighborhood Composition



The unit of analysis is the 2000 census tract envelope (N=456,926). Coefficients are estimated from the regression specified in equation 1, and have been multiplied by 100 for ease of interpretation. Additional controls include a quadratic function in neighborhood population, city and year fixed effects. Standard errors are robust and clustered to the census tract envelope level. Estimation is restricted to 2000 census tract envelopes that have a neighborhood population of at least 2,000 residents and a city population of at least 100,000. This captures 88.7 percent of neighborhoods covering 303 unique MSAs.

composition would have evolved in the same way in neighborhoods that received additional census tract splits as they would at non-split neighborhoods within the same city, absent census tract boundary updating. An F-test of the null hypothesis that the pre-period coefficients found in Figure 2 are jointly equal to zero fails to reject the null in both panels (panel (a): $F(2, 47, 187) = 1.46$, $p = 0.23$; panel (b): $F(2, 47, 187) = 2.10$, $p = 0.12$). While Figure 2 lacks evidence of differential neighborhood racial compositions prior to boundary updates, or parallel pre-trends, like all DID designs, I cannot directly test for differences in pre-trends in *unobserved* determinants of neighborhood racial composition.

The evidence provided in Figure 2 suggests that census tract delineation impacts the racial composition of neighborhoods immediately following the 2010 Census. Relative to census tract envelopes that don't split in the 2010 Decennial Census, 2000 Decennial Census envelopes that split increase in their proportion Black by 0.09 percentage points immediately following the 2010 delineation process. The effect on the envelope percent Black persists, and increases to 0.38 percentage points by 2016. At the same time, relative to census tract envelopes that don't split, 2000 Census envelopes that split decrease in their proportion white by 0.09 percentage points immediately after 2010 boundary updates and continue to decrease to 0.11 percentage points by 2016. While this effect remains persistently negative over time, standard errors are quite large beyond this first year. The traditional two period DID analysis can be found in Table A.1.

To study heterogeneous impacts of census tract splits, I examine whether the number of resulting census tracts within a 2000 Decennial Census census tract envelope differentially affect the racial composition of the neighborhood. As the number of resulting tracts increases within a given census tract envelope, there may exist additional opportunities to concentrate subsidized housing. For instance, three nested tracts resulting from the 2010 Decennial Census, within a 2000 Census envelope, suggests there could be three distinct neighborhoods eligible, instead of a single neighborhood in the case of a tract that doesn't split. To study the reduced-form impact of additional splits on neighborhood racial composition, I estimate an altered specification of the DID analysis found above:

$$y_{ict} = \gamma_0 + \gamma_1 post_t + \gamma_2 number\,of\,splits_i + \gamma_3(post_t \times number\,of\,splits_i) + x'_{it}\theta + \epsilon_{ict} \quad (2)$$

where the dependent variable y_{ict} is either the percent white or Black in census tract envelope i in city c at time t, $post_t$ indicates whether y is measured after the 2010 Decennial Census, and $number\,of\,splits_i$ indicates the number of resulting census tracts in the 2010 Decennial Census nested within 2000 Decennial Census envelope i. Additional controls in x_{it} include a quadratic function in population, city fixed effects, and a linear time trend. Standard errors are robust and clustered at the 2000 census tract envelope level.

Table 1: Heterogeneous Effects of Census Tract Splits on Neighborhood Composition

	(1) % Black	(2) % White
Number of Splits	-1.128***	1.882***
	(0.068)	(0.080)
Post × Number of Splits	0.042**	-0.036
	(0.017)	(0.023)
R^2	0.228	0.261
N	456922	456922

The unit of analysis is the 2000 census tract envelope (N=456,922). Coefficients are estimated from the regression specified in equation 2, and have been multiplied by 100 for ease of interpretation. Additional controls include a quadratic function in neighborhood population, city fixed effects, and a linear time trend. Standard errors are robust and clustered to the census tract level. Estimation is restricted to 2000 census tract envelopes that have a neighborhood population of at least 2,000 residents and a city population of at least 100,000. This captures 88.7 percent of neighborhoods covering 303 unique MSAs.

Table 1 presents the effects of the number of splits within a 2000 census tract envelope on the racial composition of the neighborhood. Relative to census tract envelopes that do not split in the 2010 Decennial Census, an additional census tract within the 2000 Census envelope results in the neighborhood Black residency increasing by 0.042 percentage points. Conditional on an 2000 Census envelope splitting, the average number of resulting 2010 census tracts nested within are 2.58 (std. dev.=1). Again, while insignificant, the coefficient of interest differs in sign when comparing proportions Black and white in census tract envelopes.

4.2 Event Study Approach

Similar to the analysis above, this section estimates the long-run impact of boundary updates using the census tract envelope approach with envelopes now fixed to their original boundaries. Using decadal data from the Longitudinal Tract Data Base (LTDB) and an event study design, I corroborate the evidence found in Section 4.1.

If available, census tract envelopes are based upon the original configuration, when a majority of census tracts were delineated for urban areas.[14] Otherwise, the envelope is defined as the first time a new census tract appears, and is not encompassed within an envelope already.[15] Additionally, the event study exercise accounts for multiple splits of the same census tract envelope.[16] In event time, this exercise captures 30 years prior to 30 years after a census tract splits. For instance, a census tract envelope that existed in 1980, but did not split until 2010 would be reflected in the relative year -30. Alternatively, a census tract envelope that was split in 1980, and observed in 2010 would be reflected as 30 years after the event. Because census tracts that were first split in the 1980 Decennial Census appear more frequently in the data while those that were split in later Censuses appear less, the final sample consists of an unbalanced panel of 38,242 census tract envelopes within the MSAs of interest.

I adopt a two-way fixed-effects (TWFE) specification to estimate the long-run impacts of boundary delineation. My estimating equation of interest is:

$$y_{ict} = \alpha_i + \alpha_t + \sum_{k=-30}^{-20} \gamma_k treatment_{it} + \sum_{k=0}^{30} \gamma_k treatment_{it} + x'_{it}\theta + \epsilon_{ict} \tag{3}$$

where the dependent variable y_{ict} is either the percent white or Black in census tract envelope i in city c at time t, and α_i and α_t are census tract envelope and time fixed effects respectively. The independent variable *treatment* equals 1 if census tract envelope i receives further delineation in year t. The base year of analysis is $k = -10$, the observed period before the census tract is split, and is thus omitted. Additional controls in x_{it} include a quadratic function in neighborhood population and city fixed effects. Standard errors are robust and clustered at the census tract envelope level.

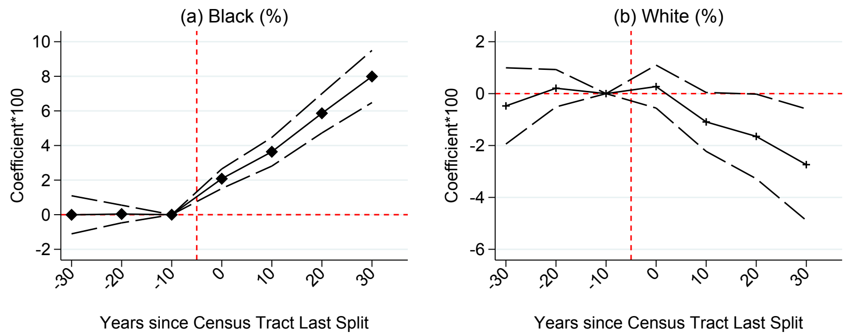
The coefficient of interest is γ_k and indicates the percentage point effect on the neighborhood racial composition in relative year k. Similar to the DID analysis above, the key identifying assumption is that census tracts would have continued to evolve in the same way if they had not received further boundary

[14] New census tracts falling within the original envelope are tagged with the number of the original envelope and appended with a decimal. To aggregate data to their original envelopes, I drop decimal points and collapse data to the original census tract envelope number.

[15] For example, as cities expanded, the Census Bureau had to further delineate this growth. This occurred up to the 1980 Decennial Census, where the Census Bureau expanded census tracts to cover the entire U.S.

[16] This occurs in the most dense areas of cities.

Figure 3: Long-Term Effect of Census Tract Splits on Neighborhood Composition



The unit of analysis is the original census tract envelope (N=38,242). Coefficients are estimated from the regression specified in equation 3. Additional controls include a quadratic function in neighborhood population and city fixed effects. Standard errors are robust and clustered to the census tract envelope level. Estimation is restricted to census tract envelopes that have a neighborhood population of at least 1,200 residents.

delineation in the following Decennial Censuses. While I cannot test this assumption directly, an F-test of the null hypothesis that the pre-period coefficients are jointly equal to zero fails to reject the null in both panels (panel (a): $F(2, 2, 758) = 0.01$, $p = 0.99$; panel (b): $F(2, 2, 758) = 1.16$, $p = 0.31$). Figure 3 presents the results of the event study analysis.

The evidence found in Figure 3 supports the results found in the short-run analysis of Figure 2. Results suggest that census tract boundary delineation significantly impacts the racial composition of neighborhoods, and effects not only persist decades but amplify over time. Panel (a) finds that ten years after a neighborhood receives further delineation, the proportion of Black residents increases 3.20 percentage points, and further increases to 5.03 percentage points twenty years after the census tract envelope is “split”, off a sample mean of 9.29 percent. Alternatively, ten years after the neighborhood is “split” the proportion of white residents decreases by 1.09 percentage points, and continues to decline thirty years after delineation. Summarized results of this exercise can be found in Table A.2

5 Mechanism Analysis: Investigating the LIHTC

The goal of this section is to consider whether federally subsidized low-income, affordable housing plays a significant role in contributing to neighborhood racial inequality through census tract delineation. While census tract boundaries are used by varying state, local, and private agencies to direct public policy, the largest federal program to do so is the Low-Income Housing Tax Credit (LIHTC) Qualified Census Tracts program. To explore this mechanism, I merge tract-aggregated LIHTC data available through the Department of Housing and Urban Development (HUD) with census tract data used earlier in the analysis to study whether census tract "splits" affect the probability of neighborhoods receiving LIHTC projects, and their associated housing characteristics. Specific outcomes of interest include heterogeneous impacts of census tract "splits", designated low-income restricted housing units, whether units are more likely to be rehabilitated or newly constructed, and neighborhood rents and housing prices.

I begin by examining whether low-income, qualified neighborhoods that receive further delineation are more or less likely to receive LIHTC-funded housing projects and how the number of resulting tracts impacts this probability. I estimate a variant of Equation 1 where the dependent variable is replaced with an indicator if the 2000 census tract envelope received an LIHTC-funded project. Next, I examine the probability of the census tract envelope receiving LIHTC funds by the number of splits within the tract envelope. To study these heterogeneous impacts, I adopt the following econometric framework:

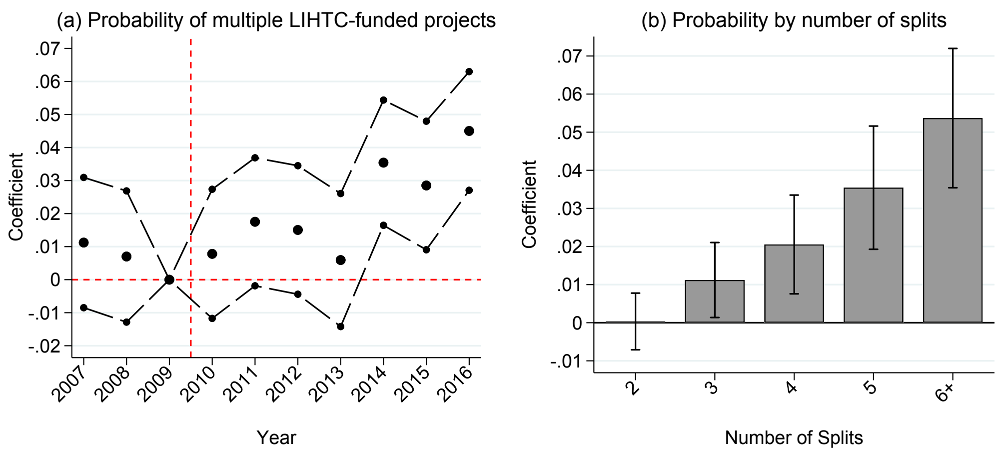
$$y_{ict} = \gamma_0 + \gamma_1 post_t + \sum_{l=2}^{6} \gamma_l number\,of\,splits_i + (post_t \times \sum_{s=2}^{6} \gamma_s number\,of\,splits_i) + x'_{it}\theta + \epsilon_{ict} \quad (4)$$

where y is an indicator if census tract envelope i received LIHTC funds, *post* indicates that time t is after the 2010 Decennial Census, and $number\,of\,splits$ indicates the number of times census tract envelope i received additional boundaries. Additional controls in x_{it} include a quadratic function in neighborhood population and city fixed effects. Standard errors are robust and clustered at the 2000 census tract envelope level. This specification allows me to test whether there exists differential impacts for census tract envelopes that are split once relative to envelopes with multiple divisions.[17] Thus, the coefficient of interest is γ_s and indicates the percentage point effect on the probability that a low-income, qualified neighborhood receives multiple LIHTC-funded projects.

The results in Figure 4 suggest that census tract delineation impacts the probability that neighborhoods receive multiple LIHTC-funded projects. Relative to census tract envelopes that don't split, 2000 qualified census tract envelopes that receive further delineation are more likely to receive multiple LIHTC-funded projects. Five years after the 2010 Decennial Census, the probability of multiple LIHTC projects within a qualified census tract envelope increases by almost 3 percentage points, off a sample mean of 64.32 percent. Panel (b) presents the heterogeneous impacts of census tract splits estimated from Equation 4, and suggests that increasing the number of resulting tracts within a 2000 qualified census tract envelope increases the probability of multiple LIHTC-funded projects. A 2000 tract envelope that is further delineated into six or more resulting census tracts after the 2010 Decennial Census increases the probability of multiple LIHTC-

[17] My sample consists of census tract envelopes that consist of up to 12 nested census tracts after the 2010 Decennial Census, but I limit the presentation of results in this exercise to envelopes that have 6 or fewer nested 2010 census tracts due to small sample sizes. Those with 6 splits or less make up over 99 percent of the analysis sample.

Figure 4: Evidence of the LIHTC in Census Tract Delineation



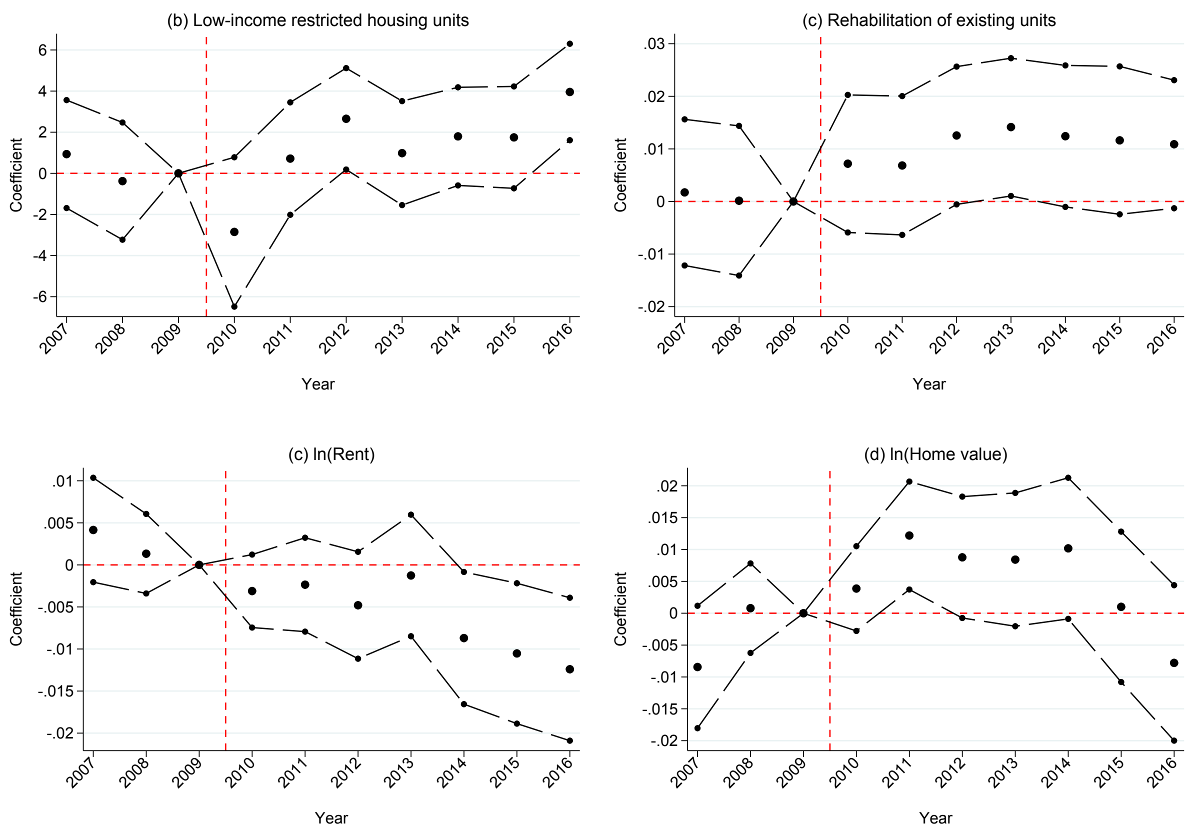
The unit of analysis is the 2000 census tract envelope and is conditional on the envelope being a low-income, qualified neighborhood from 2007 to 2016 (N=71,708). Coefficients are estimated from regressions specified in Equations 1 (Panel (a)) and 4 (Panel (b)). Additional controls include a quadratic function in neighborhood population, city fixed effects, and a linear time trend. The dependent variable of interest in each panel is an indicator that equals 1 if the census tract envelope received more than 1 LIHTC-funded project. Standard errors are robust and clustered to the census tract level.

funded projects by more than 5 percentage points. Summarized results of Figure 4 can be found in Table A.3.

Next I turn my attention to more detailed characteristics of LIHTC-funded projects, as well as associated housing market characteristics including rent and housing prices. I use a similar econometric framework as in Equation 1 where the dependent variable is replaced with the LIHTC outcome of interest and the analysis is restricted to low-income qualifying neighborhoods. The work of Dillman et al. (2017) suggests that LIHTC-funded properties within low-income neighborhoods may improve the surrounding the property values by removing blight and vacant lots in exchange for affordable housing options within distressed neighborhoods. The results found in Figure 5 support their findings.

Figure 5 suggests that census tract delineation is also related to characteristics associated with LIHTC-funded projects. While not significant at conventional levels, low-income census tract envelopes receiving further delineation are more likely to contain housing classified as low-income restricted units, and further housing market evidence supports this finding – neighborhoods that receive further delineation experience modest rent declines relative to tract envelopes that remain intact. Additionally, while further delineated neighborhoods are more likely to experience the rehabilitation of existing housing units, these improvements appear to also temporarily improve local housing values in low-income neighborhoods.

Figure 5: Census Tract Delineation and LIHTC-Funded Projects



The unit of analysis is the 2000 census tract envelope and is conditional on the envelope being a low-income, qualified neighborhood from 2007 to 2016 (N=71,708). Coefficients are estimated from the regression specified in Equation 1. Additional controls include a quadratic function in neighborhood population, city fixed effects, and a linear time trend. Low-income restricted housing units refer to units designated as separate living quarters and restricted to individuals below an income threshold determined by the state Housing Finance Agency (HFA). Rehabilitation is an indicator that equals 1 if the tract envelope received LIHTC-funded work to rehabilitate existing housing units rather than to build new structures. Median rent and home value are in 2010 dollars. Standard errors are robust and clustered to the census tract level.

6 Conclusion

Census tracts are most often applied in research to construct geographically stable neighborhoods, but a lesser known feature of census tract geography is their role in public policy. The largest and longest-running federal program to encourage the creation of affordable housing, the LIHTC, uses census tract data in order to determine low-income, qualifying neighborhoods. Because of the potential implications census tract boundaries may have throughout cities on affordable housing options, incentives to delineate strategically may arise.

In this paper, I exploit variation in census tract boundaries to describe how neighborhood dynamics, in terms of racial composition, have changed over both short- and long-run time horizons. To evaluate short-run impacts, I use a difference-in-differences empirical design surrounding the 2010 Decennial Census. Specifically, I compare the annual racial composition of census tracts that split relative to those that don't, before and after the 2010 Decennial Census holding census tract boundaries fixed to their 2000 Decennial Census "envelopes". I further corroborate my results within the short-run analysis using an alternative data set spanning 1980 to 2010 to capture long-run impacts. Lastly, I provide evidence that census tract delineation may yield such results through the Low-Income Housing Tax Credit by concentrating affordable housing in disadvantaged neighborhoods.

My first key result suggests that census tract boundaries are delineated in ways that concentrate Black residents in neighborhoods with additional census tract splits. Common trends analysis suggests this can not be explained by observable differences prior to the 2010 Decennial Census. Further analysis using long-run data, I find that these effects persist decades and amplify over time – twenty years beyond census tract delineation results in the neighborhood proportion of Black residents increasing 5.03 percentage points, off a pre-treatment sample mean of 9.29 percent. I further provide evidence that the LIHTC may play a role in concentrating Black residents in neighborhoods with more tract boundaries. Conditional on low-income, qualifying neighborhoods, 2000 census tract envelopes are 1.6 percentage points more likely to receive multiple projects relative to neighborhoods that did not receive further delineation.

To my knowledge, this is the first paper to consider the impacts that census tract delineation may have on the development and dynamics of urbanizing neighborhoods. A significant literature has documented the disparities and inequalities that result from unequal access to opportunity across neighborhoods, and this paper explores a new channel that may contribute to increasing neighborhood inequality. To the extent that public policy, such as the LIHTC program, relies on the delineation of census tracts in determining qualified neighborhoods, little work exists in evaluating the consequences of such criteria. Future work should continue to explore avenues in which neighborhood boundaries may influence the composition and direction of public policy.

References

Aaronson, Daniel, Daniel Hartly, and Bhashkar Mazumder. Forthcoming. "The effects of the 1930s HOLC "redlining" maps." *American Economic Journal: Economic Policy.*

Ananat, Elizabeth Oltmans. 2011. "The wrong side (s) of the tracks: The causal effects of racial segregation on urban poverty and inequality." *American Economic Journal: Applied Economics* 3 (2): 34–66.

Bayer, Patrick, and Robert McMillan. 2012. "Tiebout sorting and neighborhood stratification." *Journal of Public Economics* 96 (11-12): 1129–1143.

Bayer, Patrick, Robert McMillan, Alvin Murphy, and Christopher Timmins. 2016. "A dynamic model of demand for houses and neighborhoods." *Econometrica* 84 (3): 893–942.

Card, David, Alexandre Mas, and Jesse Rothstein. 2008. "Tipping and the Dynamics of Segregation." *The Quarterly Journal of Economics* 123 (1): 177–218.

Chetty, Raj, and Nathaniel Hendren. 2018. "The impacts of neighborhoods on intergenerational mobility II: County-level estimates." *The Quarterly Journal of Economics* 133 (3): 1163–1228.

Chetty, Raj, Nathaniel Hendren, Maggie R Jones, and Sonya R Porter. 2020. "Race and economic opportunity in the United States: An intergenerational perspective." *The Quarterly Journal of Economics* 135 (2): 711–783.

Chetty, Raj, Nathaniel Hendren, and Lawrence F Katz. 2016. "The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment." *American Economic Review* 106 (4): 855–902.

Collins, William J., and Robert A. Margo. 2000. "Residential segregation and socioeconomic outcomes: When did ghettos go bad?" *Economics Letters* 69 (2): 239–243.

Cutler, David M., and Edward L. Glaeser. 1997. "Are ghettos good or bad?" *The Quarterly Journal of Economics* 112 (3): 827–872.

Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor. 2008. "When are ghettos bad? Lessons from immigrant segregation in the United States." *Journal of Urban Economics* 63 (3): 759–774.

Dawkins, Casey. 2013. "The spatial pattern of low income housing tax credit properties: Implications for fair housing and poverty deconcentration policies." *Journal of the American Planning Association* 79 (3): 222–234.

Dillman, Keri-Nicole, Keren Mertens Horn, and Ann Verrilli. 2017. "The what, where, and when of place-based housing policy′s neighborhood effects." *Housing Policy Debate* 27 (2): 282–305.

Ellen, Ingrid Gould, and Keren Mertens Horn. 2018. "Points for place: Can state governments shape siting patterns of low-income housing tax credit developments?" *Housing Policy Debate* 28 (5): 727–745.

Farrell, Chad R, and Barrett A Lee. 2011. "Racial diversity and change in metropolitan neighborhoods." *Social Science Research* 40 (4): 1108–1123.

Fischer, Will. 2018. "Low-Income Housing Tax Credit Could Do More to Expand Opportunity for Poor Families." *Washington, DC: Center on Budget and Policy Priorities.*

Galiani, Sebastian, Alvin Murphy, and Juan Pantano. 2015. "Estimating neighborhood choice models: Lessons from a housing assistance experiment." *American Economic Review* 105 (11): 3385–3415.

Hollar, Michael, and Kurt Usowski. 2007. "Low-income housing tax credit qualified census tracts." *Cityscape* 153–159.

Jubara, Amalea, Jonathan Schroeder, and Yaxuan Zhang. 2021. "Mapping block-level segregation: The twin cities' black population, 1980-2010." IPUMS Blog, February.

Kennedy, Anthony. 2015. "Texas department of housing & community affairs v. the inclusive communities project, inc." *Opinion of the United States Supreme Court* 6.

Logan, John R., Brian Stults, and Zengwang Xu. 2014. "Longitudinal Tract Database [dataset]." http://www.s4.brown.edu/us2010/Researcher/Bridging.htm.

Manson, Steven, Jonathan Schroeder, David Van Riper, and Steven Ruggles. 2019. "IPUMS National Historical Geographic Information System: Version 14.0 [Database]." Minneapolis, MN: IPUMS.

Nagaraj, Abhishek, and Scott Stern. 2020. "The Economics of Maps." *Journal of Economic Perspectives* 34 (1): 196–221.

Office of Policy Development and Research. 2013. "HUD Desingates Low-Income Housing Tax Credit Qualified Census Tracts for 2013."Technical report, U.S. Department of Housing and Urban Development.

Owens, Ann. 2016. "Inequality in children's contexts: Income segregation of households with and without children." *American Sociological Review* 81 (3): 549–574.

Reid, Carolina K. 2019. "Rethinking opportunity in the siting of affordable housing in California: Resident perspectives on the low-income housing tax credit." *Housing Policy Debate* 29 (4): 645–669.

Scally, Corianne, Amanda Gold, and Nicole DuBois. 2018. "The Low-Income Housing Tax Credit: How it Works and Who It Serves."Technical report, Urban Institute.

Sharkey, Patrick. 2016. "Neighborhoods, cities, and economic mobility." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 2 (2): 159–177.

Tax Policy Center. 2017. "Tax Policy Center Briefing Book." *Tax Policy Center*.

U.S. Census Bureau. 1994. *Geographic areas reference manual.* US Dept. of Commerce, Economics and Statistics Administration, Bureau of the Census.

U.S. Census Bureau. 2002. "Measuring America: The Decennial Censuses from 1790 to 2000." Retrieved from https://www.census.gov/history/pdf/measuringamerica.pdf.

U.S. Census Bureau. 2008. "2010 Participant Statistical Areas Program Participant Information."Technical report, U.S. Department of Commerce.

U.S. Census Bureau. 2011. "Historical Delineation Files." Retrieved from https://www.census.gov/geographies/reference-files/time-series/demo/metro-micro/historical-delineation-files.html.

U.S. Census Bureau. 2013. "Census Tract Education." Retrieved from https://www2.census.gov/geo/pdfs/education/CensusTracts.pdf.

U.S. Const. art. I, § 2.. 1787. "Constitution of the United States."

Appendix A

Table A.1: DID Effects of Census Tract Splits on Neighborhood Composition

	(1) % Black	(2) % White
Treatment	-2.644***	4.164***
	(0.206)	(0.234)
Post × Treatment	0.262***	-0.164**
	(0.050)	(0.067)
Mean of dep. variable	14.21%	70.94%
R^2	0.228	0.259
N	456926	456926

The unit of analysis is the 2000 census tract envelope (N=456,926). Coefficients are estimated from the following regression: $y_{ict} = \beta_0 + \beta_1 post_t + \beta_2 treatment_i + \beta_3(post_t \times treatment_i) + x'_{it}\theta + \epsilon_{ict}$, and have been multiplied by 100 for ease of interpretation. The variable $post_t$ indicates whether y is measured after the 2010 Decennial Census, and $treatment_i$ indicates whether census tract envelope i receives further delineation in the 2010 Census. Additional controls include a quadratic function in neighborhood population, city fixed effects, and a linear time trend. Standard errors are robust and clustered to the census tract level. Estimation is restricted to 2000 census tract envelopes that have a neighborhood population of at least 2,000 residents and a city population of at least 100,000. This captures 88.7 percent of neighborhoods covering 303 unique MSAs.

Table A.2: Long-Run Impacts of Census Tract Splits on Neighborhood Composition

	(1) % Black	(2) % White
Treatment	1.818***	-0.326
	(0.344)	(0.444)
Mean of dep. variable	9.29%	75.92%
R^2	0.457	0.642
N	38,242	38,242

The unit of analysis is the original census tract envelope (N=38,242). Coefficients are estimated from the regression specified in equation 3. Treatment is an indicator that equals 1 if the unit is observed in the relative year $0 \leq k \leq 30$, and thus reflects the average percentage point effect over this period. Additional controls include a quadratic function in neighborhood population and city fixed effects. Standard errors are robust and clustered to the census tract envelope level. Estimation is restricted to census tract envelopes that have a neighborhood population of at least 1,200 residents.

Table A.3: Evidence of the LIHTC in Census Tract Delineation

	(1) Multiple LIHTC-Funded Projects	(2) Multiple LIHTC-Funded Projects
Treatment	-0.069***	
	(0.004)	
Post × Treatment	0.016***	
	(0.005)	
Number of splits		-0.024***
		(0.001)
Post × Number of splits		0.006***
		(0.001)
Mean of dep. var.	64.32%	64.32%
R^2	0.021	0.022
N	71708	71708

The unit of analysis is the 2000 census tract envelope and is conditional on the envelope ever receiving an LIHTC-funded project from 2007 to 2016 (N=71,708). Coefficients are estimated from regressions specified in equations X and Y. Additional controls include a quadratic function in neighborhood population, city fixed effects, and a linear time trend. The dependent variable of interest in each column is an indicator that equals 1 if the census tract envelope received more than 1 LIHTC-funded project. Standard errors are robust and clustered to the census tract level.