Accounting for Central Neighborhood Change, 1980-2010

Nathaniel Baum-Snow, Daniel Hartley*

October 21, 2017

Abstract

Central neighborhoods of most U.S. metropolitan areas experienced 1980-2000 population declines and 2000-2010 revitalization. 1980-2000 departures of residents without a college degree accounted for most of the declines while the return of college educated whites and the stabilization of neighborhood choices by less educated whites drove most of the post-2000 rebounds. Increases in amenity valuations after 2000 encouraged college-educated whites to move in and other whites to remain. Continued departures of less than college educated minorities were mainly driven by their continued reductions in demand for downtown amenities.

*Baum-Snow: University of Toronto, 105 St. George St. Toronto, ON M5S 3E6, Nate.Baum-Snow@rotman.utoronto.ca; Hartley: Federal Reserve Bank of Chicago, 230 S. LaSalle St., Chicago, IL 60604, Daniel.A.Hartley@chi.frb.org. We thank participants in various seminars and conferences for their helpful comments. The views expressed are those of the authors and do not necessarily represent the views of the Federal Reserve Bank of Chicago, the Board of Governors of the Federal Reserve System, or their staff.

1 Introduction

In the decades following WWII, the central regions of most U.S. metropolitan areas were in decline. Between 1960 and 2000, the aggregate central city population share in the 100 largest metropolitan areas fell from 0.49 to 0.24 while the employment share declined from 0.61 to 0.34 (Baum-Snow, 2017). A host of mechanisms responsible for this decline have been considered in the literature. These include highway construction (Baum-Snow, 2007), decentralization of low-skilled work (Kain, 1992), white flight from rising minority populations in cities (Boustan, 2010), rising incomes (Margo, 1992), Federal Housing Authority mortgage insurance provision favoring the suburbs (Jackson, 1985) and vintage housing in cities filtering down to lower-income occupants (Brueckner and Rosenthal, 2009). Following sharp population and economic declines during the 1970s, neighborhoods within 2 km of central business districts (CBDs) in the largest U.S. metropolitan areas experienced continued 1980-2000 declines in population, averaging 7 percent. However, population, income and college fraction all grew on average in these central neighborhoods during the 2000-2010 period. Though 2000-2010 population growth within 2 km of CBDs averaged 6 percent, almost equal to the aggregate growth rate in the sample, downtown neighborhoods were among the most rapidly gentrifying regions of metropolitan areas during the 2000-2010 period when measured in terms of fraction white, income and fraction with a college degree. This paper investigates the factors that drove this 1980-2000 decline and 2000-2010 gentrification of the central neighborhoods of large U.S. metropolitan areas.

Our evaluation of the causes of central neighborhood change proceeds in two stages. First, using a procedure akin to that proposed by DiNardo, Fortin & Lemieux (1996) for decomposing wage distributions, we systematically decompose the sources of changes in demand for central neighborhoods since 1980 into those due to secular demographic shifts holding neighborhood choices constant and those due to changes in neighborhood choices of particular demographic groups holding demographic shares constant. We carry out the analysis using cells defined by joint population distributions of race and education, age, family structure, or household income decile. While our focus is on central neighborhoods, this methodology can be applied more broadly to decompose the drivers of change for any type of neighborhood.

Second, to better understand why groups' neighborhood choices changed, we use a conditional choice probability procedure to recover changes in valuations of various neighborhood attributes in each decade from 1980 through 2010 in the context of a neighborhood choice model. The model shows how to combine information about observed neighborhood choices and housing costs to recover neighborhood valuations that reflect a combination of sub-metropolitan area labor market opportunities and local amenities. Using model estimates, we evaluate the extent to which shifts in housing costs, labor demand conditions and various components of demand for local amenities by different groups promoted 1980-2000 and 2000-2010 changes in central neighborhood population and demographic composition.

The central result from the decompositions is that most of central neighborhood change has been driven by the fact that whites have chosen to live in CBD area neighborhoods at much higher rates after 2000 relative to the prior decades, with this phenomenon particularly strong for more educated and higher income whites. Indeed, the area within 2 km is the only CBD distance ring within 20 km in which the white population grew over the 2000-2010 period on average across CBSAs, with 1980-2000 departures of low socioeconomic status (SES) minorities continuing unabated. Shifts in neighborhood choices drove 1980-2000 central neighborhood population decline despite the fact that growth in minority share would bolster demand for these neighborhoods holding neighborhood choices constant. Low SES nonwhites' declines in central neighborhood choice probabilities in each decade over the full 1980-2010 study period represents the largest force for population declines. High SES whites' slight declines in 1980-2000 central neighborhood choice probabilities reversed after 2000 to generate the majority of central area population growth. However, the main driver of the turnaround of central neighborhoods comes from the fact that low-SES whites stopped departing central neighborhoods after 2000. Changing neighborhood choices of high-SES minorities had only small impacts. More rapid 1980-2000 departures of low-SES households from central neighborhoods contributed to growth of average incomes in these neighborhoods, even in the face of declining populations.

Since central area residents are disproportionately minority, the growing share of minorities in the U.S. population have consistently pushed in favor of downtown population growth. Indeed, without this force central neighborhoods would have experienced continued population declines after 2000. Shifts in the distribution of family structure (the growing share of households without children) conditional on race have pushed in favor of population growth as well since 1980, making it unlikely that these changes have driven the reversal of downtown population declines. Shifts in the income distribution and the age structure of the population conditional on race have also had small effects.

To recover mechanisms driving shifts in neighborhood choices, we develop a model that incorporates insights from Berry (1994) and Bayer et al. (2016) and facilitates recovery of the relative importance of changing labor market opportunities, housing costs and amenities for driving each group's shifts in neighborhood choices. Our estimates indicate that while each group responded to improved central area labor market opportunities about equally, different income elasticities of demand for downtown amenities emerged across groups after 2000. While in the 1980s, income growth drove suburbanization of all demographic groups, consistent with Margo's (1992) evidence, 2000-2010 income growth of college educated whites no longer deterred them from living downtown. In addition, we find evidence of increasing amenity valuations of downtown neighborhoods (holding incomes constant) for all groups except low SES minorities.

Using estimates from our neighborhood choice model, we carry out unified decompositions of the mechanisms driving the components of central area population change driven by group-specific neighborhood choice probabilities. These decompositions reveals that 1980-2000 reductions in the quality of central neighborhood amenities was the most important driver of these areas' population declines except amongst low SES whites, for whom income effects on existing local amenities were most important. For the 2000-2010 period, income effects were the most important driver of continued departures of low SES minorities, while the impact of unobserved amenities turned positive for other groups. While all groups value improved downtown labor market opportunities, the average CBSA experienced declining downtown employment in the 1990s and essentially no change in downtown employment in the 2000-2010 period. As a result, shifts in central area labor market opportunities had a miminal impact on central area population changes since 2000, though the stabilization of downtown employment declines represents a force that promoted stabilization after 1980-2000 downtown population declines.

Our conclusion that shifts in amenity valuations rather than in nearby labor market opportunities or housing cost changes have primarily driven changes in central neighborhood choices echoes evidence in Couture & Handbury (2016), which performs a detailed investigation of which amenities are driving these shifts for the young and college educated. However, our evidence that all groups' central neighborhood valuations are increasing in nearby labor market opportunities are also in line with those in Edlund, Machado, & Sviatchi (2015), though that paper focuses on larger cities with more robust 2000-2010 employment growth than is seen in our broader sample. Our finding of important racial differences in trends in the valuation of downtown amenities, especially amongst low SES individuals and households, is an important part of the broader narrative about central neighborhood change that has not been considered elsewhere in the literature.

2 Characterizing Neighborhood Change

In this section, we establish facts about central neighborhood change between 1970 and 2010 that represent a baseline for the analysis in subsequent sections.

2.1 Data

We primarily use 1970-2010 decennial Census data and the 2008-2012 American Community Survey (ACS) data tabulated to the 2000 definition of census tract boundaries for the analysis. Central to our investigation is the need for joint distributions of population by race, education, household income, age and family structure across census tracts in each CBSA. To recover as many of these joint distributions in the most disaggregated form possible, we make use of both summary tape file (STF) 3 and 4 census tabulations. We also use information about family structure and age by race from STF1 data from the 2010 Census. Because the 2010 Census did not collect information about income or education, we must rely on the 5-year ACS data for these tract distributions. We also make use of some census micro data to estimate parameters governing shapes of household income distributions above topcodes, to generate weights used to assign some of the population counts in the tract aggregate data to different types of families, and to estimate housing expenditure shares by household demographic type. All census tracts are normalized to year 2000 geographies using allocation factors from the U.S. Census Bureau.

We construct three different joint distributions for people and one for households in 1980, 1990, 2000 and 2010. For each one, the race categories are white, black and other. In the other dimension of each joint distribution, we have either 4 education groups (less than high school, high school only, some college, college +), 18 age groups (0-4, 5-9, ..., 80-84, 85+) or 6 family type groups (in married couple families with no children, in married couple families with children, in single female headed families with children, in single male headed families with children, not in a family, in group quarters). For income, we construct the number of households in each decile of the household income distribution of those residing in our sample area in each year. We do this in order to facilitate comparisons across CBSAs and years in a sensible way while taking into account the secular increase in nationwide income inequality during our sample period.

The Census Transportation Planning Package (CTPP) reports aggregated census or ACS micro data to microgeographic units for place of work in 1990, 2000 and 2005-2009. We use these data broken out by industry to construct localized labor demand shocks. Where available, we take CBD definitions from the 1982 Economic Census. Otherwise, we use the CBD location as assigned by ESRI. Each CBSA is assigned only one CBD.

Our sample includes the regions of all 120 year 2008 definition metropolitan areas (CBSAs) that were tracted in 1970 and had a population of at least 250,000 except Honolulu.[1] In order for our analysis to apply for the average metropolitan area rather than the average resident, much of the analysis applies tract level weights such that each CBSA is weighted equally. So that parameters that govern demand conditions for regions within 4km of CBDs represent the average CBSA in our sample, we also weight this region equally across CBSAs. The Data Appendix provides more details about our data construction.

To achieve a succinct descriptive analysis, we construct a summary measure of neighborhood demographics that incorporates the share of residents that are white, the share that are college educated and the median household income of the tract. This summary measure for tract i is the average number of standard deviations tract i is away from its CBSA mean in each year for each of these components. We call this equally weighted tract z-score the socioeconomic status (SES) index.[2]

Figure 1 shows a map of the 120 CBSAs in our sample shaded by the fraction of census tracts within 4 km of the central business district that are in the top half of the tract distribution of our SES index in 1980 (top left) and 2010 (top right) in each CBSA. Those CBSAs above 0.5 have central areas that are less distressed than would be expected given a random assignment of SES status to census tracts. Particularly striking is the number of CBSAs whose central areas experience gentrification between 1980 and 2010 (moving up the distribution of blue-red shades). Santa Barbara and New York are the only CBSAs with downtown areas that were more affluent than average in 1980. By 2010, 9 additional CBSAs had relatively affluent downtown areas. While central areas of other CBSAs remained less affluent than average, most became more affluent between 1980 and 2010. Of the 120 CBSAs in our sample, the fraction of the population within 4 km of a CBD living in a tract in the top half of the SES index distribution increased by more than 0.25 in 15 CBSAs, by 0.10 to 0.25 in 35 CBSAs and by 0.00 to 0.10 in 23 CBSAs between 1980 and 2010. Central areas of the remaining 47 CBSA experienced only small declines in their SES indexes on average. These patterns

[1] 100% of the 2000 definition tract must have been tracted in 1970 to be in our sample.

[2] While race is not a measure of socioeconomic status, there is evidence that conditional on income and education, black households have lower wealth than white households (Altonji, Doraszelski, and Segal, 2000). We include the share of residents that are white in our SES index as a proxy for unobserved elements of socioeconomic status such as wealth.

of changes are seen at the bottom of Figure 1, with red shaded CBSAs experiencing a rise in SES in central areas and the blue shaded CBSAs a decline in SES in central areas.

2.2 Facts About Neighborhood Change

Figure 2 reports statistics describing various aspects of neighborhood change as functions of the distance from the CBD since 1970. All plots show medians across CBSAs in our sample. We choose medians in order to emphasize that changes are not driven by just a few large notable cities. Analogous results using means across CBSAs or aggregates are similar. The broad message from Figure 2 is that downtown gentrification since 2000 is evident in many dimensions and is very localized. Neighborhoods within 2 km of CBDs experienced the fastest 2000-2010 growth in terms of population, the share of residents that are white, and the share of residents that are college-educated of all CBD distance bands. The seeds of this gentrification started to form after 1980, as evidenced by more localized upticks in these indicators right at CBDs.

The evidence in Figure 2 shows that while central area population growth relative to that in the suburbs is a useful indicator of downtown gentrification, two additional features in the data also indicate a turnaround in overall demand for downtown neighborhoods. First, the growth in population growth (the second derivative) is positive well beyond 2 km from the CBD. At each distance out to 10 km, the population growth rate increased in the city relative to at 20 km from the CBD in each decade after the 1970s, with this relative increase roughly monotonically declining as a function of CBD distance in the 1980s and the 2000-2010 periods. Second, even areas within about 5 km of the CBD that experienced declining 2000-2010 populations experienced faster than average growth in fraction white and fraction college educated. As we show below in the context of our neighborhood choice model, these types of residental composition shifts represent increasing aggregate demand for living in these central neighborhoods.

Table 1 reports transitions of individual census tracts through the distributions of three indicators. We present this evidence about the nature of demographic change in central neighborhoods to provide a sense of the heterogeneity around the summary statistics presented in Figure 2 and to show that a few neighborhoods moving quickly up the distribution are not driving central area gentrification. Table 1 shows the fraction of the population within 4 km of a typical CBSA's CBD living in tracts moving more than 20 percentile points or 0.5 standard deviations up or down the CBSA tract distribution. When calculating these numbers we weight by the tract's share of CBSA population in the base year, meaning all CBSAs are weighted equally. Commensurate with the evidence in Figure 2, two of the three measures indicate that central area tracts were, on balance, in decline during the 1970s, with these declines slowly reversing sometime in the 1980s or 1990s. As in Figure 2, evidence in Table 1 shows that the resurgence of the central areas really took off between 2000 and 2010.[3]

To help visualize typical trends in neighborhood inequality at the CBSA level, Figure 3 depicts the same three measures of neighborhood change in the Chicago CBSA between 1980 and 2010. We calculate demeaned share white (Panel A), demeaned college-graduate share (Panel B) and demeaned percentile ranking of the tract's median household income within our sample of tracts (Panel C). We calculate these measures for each tract in 1980 and 2010, weighting by tract population. These indicators are graphed against each other in a scatterplot, with 45-degree and regression lines indicated. Both of these lines pass through (0,0) in each panel by construction. Dark black dots represent tracts within 4 km of the CBD. Regression slopes of less than 1, for tract income percentile and tract share white indicate that Chicago neighborhoods have experienced convergence in these dimensions. Points above a regression line that are far to the left of a 1980 mean represent gentrifying census tracts.

Figure 3 reveals considerable heterogeneity in Chicago neighborhood change over the period 1980-2010, with our three neighborhood change measures clearly capturing distinct things. The masses of points at the bottom left and top right of Panel A represent large concentrations of stable minority and white census tracts, respectively. The relatively large number of tracts along the right edge of the graph at almost 100 percent white in 1980 and ending up less than 70 percent white may have experienced tipping (Card, Mas & Rothstein, 2008). But a handful of tracts went in the other direction between 1980 and 2010, seen in the upper left area of the graph. These largely minority tracts in 1980, that gained white share much faster than the typical Chicago tract, are almost exclusively within 4 km of the CBD. Indeed, all but 4 of the tracts within 4 km of the CBD that were less than 80 percent white in 1980 experienced increases in white share between 1980 and 2010, even though share white decreased on average. Such downtown area gentrification is clear from the other measures in Figure 3 as well. Central area tracts are clustered in the upper left area of each panel.

[3] Downtown neighborhoods were the poorest and had among the lowest education levels and shares of white residents of any CBD distance ring in 1980. One potential explanation for downtown gentrification is, thus, simple mean reversion. In Section 5.1 we provide evidence that while mean reversion in neighborhood income and racial composition does exist, it is not the main force behind downtown revitalization.

3 Counterfactual Neighborhood Compositions

The results in the previous section show that central neighborhoods have been chosen at higher rates by high-SES demographic groups since 2000. Thus far, our examination of location choices one demographic group at a time has limited our ability to determine which groups' shifts in neighborhood choices have driven downtown gentrification, especially since college education, high incomes and racial composition are all strongly correlated. In addition, the analysis to this point has not evaluated the extent to which demographic change toward more education, a more unequal income distribution and smaller families has contributed to central area gentrification. To separate out the relative importance of changing race-specific neighborhood choices from other observed demographic factors that may be correlated with race, we use tract-level joint distributions of race and education or race and age, family structure or income over time to construct counterfactual neighborhood compositions absent changes in neighborhood choices for particular race-education (and race-other factor) combinations. The analysis simultaneously evaluates the extent to which population growth and SES improvement in central neighborhoods are driven by shifts in the demographic compositions of CBSA populations.

Our decompositions follow the logic developed by DiNardo, Fortin & Lemieux (1996) for decomposing wage distributions. To quantify the relative importance of changing neighborhood choices versus demographic shifts for generating neighborhood change, we calculate magnitudes of central area population and demographic change under various counterfactual scenarios. First, we hold the fraction of the CBSA population made up by various demographic groups fixed over time but allow neighborhood choices by specific groups to shift as in equilibrium, one at a time. This allows us to evaluate the extent to which changes in the choices of high-SES individuals and household by race have driven central neighborhood change while holding the demographic composition of CBSA populations constant. We then additionally calculate how shifts in the CBSA-level compositions of various demographic groups, conditional on race, have mechanically influenced neighborhood change. Finally, we quantify the impacts of CBSA-level racial change on central area population and demographics. This procedure, laid out in more detail below, has similarities to that developed in Carillo & Rothbaum (2016).

This decomposition allows us to identify distinct forces driving central neighborhood change in the 1980-2000 versus the 2000-2010 periods. In the earlier period, central neighborhoods experienced the flight of the poor, less educated and households with children. This was true for both white and minority households. Their flight was sizable enough to counterbalance a growing minority population, which mechanically increased the population of central area incumbent demographic groups. By 2000, there was a clear shift in the racial and SES makeup of near CBD neighborhoods. The movement of high-SES whites into central neighborhoods strengthened while the outflow of low-SES whites ceased or reversed. Over the entire study period, the increasing college-graduate share in the population, especially among whites, has also been important for driving composition shifts of downtown neighborhoods toward being more white and educated.

3.1 Construction of Counterfactual Neighborhoods

We observe the joint population distribution $f_{jt}(i, r, x)$ of race r and the other demographic attribute x across census tracts i in CBSA j in year t. The attribute x indexes education group, age group, family structure or household income decile in the national distribution. That is, $f_{jt}(i, r, x)$ denotes the fraction of CBSA j population at time t that is in demographic group (r, x) and lives in tract i. Given the structure of the tabulated census data, we are forced to evaluate counterfactual joint distributions of race (white, black, or other) and only one other demographic attribute at a time across census tracts. Denote N_{jt} as the total population of CBSA j at time t and CBSA density functions of demographics as $g_{jt}(r, x) = \sum_i f_{jt}(i, r, x)$. Crucially, we treat CBSA-level allocations $g_{jt}(r, x)$ and populations N_{jt} as exogenous to the allocation of people across neighborhoods, which can be justified in a long-run open-city model as in Ahlfeldt et al. (2015). Therefore, while aggregate population does not influence conclusions drawn from these mechanical counterfactuals, it will be incorporated in the model below when we consider the extent to which housing cost changes have driven neighborhood demographic change.

Central to our recovery of counterfactuals is the following decomposition:

$$f_{jt}(i, r, x) = f_{jt}(i|r, x)g_{jt}(x|r)h_{jt}(r) \tag{1}$$

This expression shows how to separate out neighborhood choices of particular demographic groups $f_{jt}(i|r, x)$ from the CBSA-level distribution of (r, x) across locations. It additionally shows how to separate out shifts in education, age, income, or family type compositions independent of racial composition. Components of demographic change driven by changes in demand by group (r, x) for tract i are captured by shifts in $f_{jt}(i|r, x)$. In this section we do not attempt to determine why neighborhood choices $f_{jt}(i|r, x)$ change, but only to isolate which groups' changes in neighborhood choices drive overall patterns in the data. Our causal analysis comes in the next section in the context of a neighborhood choice model. That is, $f_{jt}(i|r,x)$ can be viewed as an equilibrium object that may depend on choices of all other demographic groups. Components driven by changes in the demographic makeup of whites, blacks or other minorities holding the racial distribution constant are captured by shifts in $g_{jt}(x|r)$. Components driven by changes in the racial composition of the population holding the demographic makeup of each race constant are captured by shifts in $h_{jt}(r)$.

Tables 2-4 report results of the counterfactual experiments. Tables 2 and 3 separates out mechanisms driving total central area population change. Table 4 decomposes sources of changes in central areas' share white, share college-graduate and median income. In each table, Panels A and B report results for 1980-2000 and 2000-2010 respectively. Table 2 focuses on joint distributions of race and education for 2 km and 4 km CBD distance rings. In Table 3, each row uses a different data set with joint distributions of race with age, family type and income. We now walk through the construction of the counterfactual data sets that are used to construct output for each column in the tables.

Column 1 in Tables 2-4 reports changes in outcomes of interest for central areas calculated using the raw data as a basis for comparison with counterfactuals. Because of sampling variability across the education, age and family type data sets and the use of households rather than people in the income data set, the numbers in Column 1 of Tables 2 and 3 do not match perfectly across data sets. Column 2 shows the change that would have occurred had choices and shares not shifted from the base year. In Tables 2 and 3, this is the CBSA population growth rate. Because objects of interest in Table 4 are invariant to scale, Column 2 is all 0s in this table.

Remaining columns of Tables 2-4 are built using counterfactual distributions. Our notation indicates column number superscripts on these probability density functions. Column 3 of Tables 2-4 reports counterfactual central neighborhood change given CBSA demographic shares that are unchanged from the base year. In particular, they are constructed using the counterfactual distributions

$$f^3_{jt}(i,r,x) = f_{jt}(i|r,x)g_{jb}(x|r)h_{jb}(r).$$

Here, demographic shares $g_{jb}(x|r)h_{jb}(r)$ are for the base year but neighborhood choices for each group $f_{jt}(i|r,x)$ change as they did in equilibrium. Results in Column 4 of Tables 2-4 show the effects of holding choices constant but allowing demographic shares to shift as in equilibrium. These statistics are constructed using the counterfactual distribution

$$f_{jt}^{4}(i,r,x) = f_{jb}(i|r,x)g_{jt}(x|r)h_{jt}(r).$$

In most cases, baselines in Column 1 are closer to the results in Column 3 than they are to the than the results in Column 4. This means that changes in neighborhood choices have been more important than changes in neighborhood shares for generating observed patterns in the data.

Columns 5-10 in Tables 2-4 decompose the difference between the actual changes in Column 1 and the counterfactuals given no changes in choices or shares (in Column 2) into components that are related to changes in neighborhood choices (Columns 5-8) and demographic shares (Columns 9-10). The four effects in Columns 5-8 sum to the total effect of changing choices holding demographic shares constant reported in Column 3 (relative to no changes reported in Column 2). Adding the effects of changing demographic shares yields the full difference between the actual data in Column 1 and the "no changes" baseline in Column 2. That is, taking a cumulative sum from left to right starting at Column 5 can be thought of as piling on additional components of demographic change from a baseline of no changes in Column 2 to equal the full changes in Column 3.

Columns 5-8 report components of changes in equilibrium tract composition due to changing neighborhood choices of target whites, non-target whites, target non-whites and non-target non-whites, respectively, holding demographic shares at their base year levels. “Target" refers to college graduates, 20-34 year olds, single people and married couples without children, or households in the top three deciles of the income distribution of the full sample area, depending on the data set used.

In particular, the set of results for counterfactual c (5 to 8) is constructed using distributions built as

$$f_{jt}^{c}(i,r,x) = f_{jt}^{c}(i|r,x)g_{jb}(x|r)h_{jb}(r),$$

where $f_{jt}^{c}(i|r,x) = f_{jt}(i|r,x)$ for the elements of (r,x) listed in the column headers and $f_{jt}^{c}(i|r,x) = f_{jb}(i|r,x)$ for the remaining elements of (r,x). We note that the order of demographic groups for which we impose year t choices does not affect the results. This is because the change in the fraction of the population in tract i as a result of imposing any of these counterfactuals is linear. Each counterfactual amounts to imposing year t rather than year b choices for a few additional elements of (x,r) at a time. Mathematically, the difference in the fraction of the population living in tract

i associated with counterfactual c relative to $c-1$ is

$$\sum_x \sum_r [f^c_{jt}(i|r,x) - f^{c-1}_{jt}(i|r,x)]g_{jb}(x|r)h_{jb}(r). \tag{2}$$

Because of linearity within the square brackets, Equation (2) indicates that the full choice adjustment counterfactual 3 can be achieved by imposing counterfactuals 5, 6, 7 and 8 cumulatively in any order. Equation (2) also indicates that counterfactual c's influence on tract composition depends not only on the magnitudes of differences in choices made by the group (x, r) in question between t and the base year $[f^c_{jt}(i|r,x) - f_{jb}(i|r,x)]$, but also on the fraction of that group in the CBSA population in the base year, $g_{jb}(x|r)h_{jb}(r)$. That is, neighborhoods change the same amount if a large group makes small changes in neighborhood choices or a small group makes large changes in neighborhood choices. To provide information about which one is driving the results, Table 3 reports the average fraction of CBSA populations in parentheses for each of the four sets of demographic groups for which we examine the effects of changes in choices.

The key comparison that drives our calculations about the importance of changes in neighborhood choices by a particular group (r', x') is between $f_{jt}(i|r', x')$ and $f_{jb}(i|r', x')$, holding the neighborhood choice probabilities of other groups constant. We recognize that in practice neighborhood choice probabilities may be interdependent in equilibrium. As a result, a counterfactual in which choice probabilities are simultaneoulsy $f_{jt}(i|r', x')$ for group (r', x') and $f_{jb}(i|r'', x'')$ for group (r'', x''), holding overall demographic shares constant, may not be the equilibrium of a model in which all groups choose neighborhoods simultaneously. Rather than explore counterfactual equilibria, we emphasize that our main objective in this section is to perform a systematic accounting of neighborhood changes that did occur. We recognize the possibility that these changes have been driven in part by endogenous demand interactions between demographic groups. Empirical implementation of the model in the following section addresses this possibility.

After determining the roles of changes in neighborhood choices holding demographic composition constant, the remaining changes must be due to shifts in population composition. To look at this, we first maintain the base year racial distribution and examine how shifts in other demographic attributes conditional on race have influenced neighborhood choices. This allows us to see the influences that rising education levels, changes in income inequality, more single people, and the aging of the population have had on downtown neighborhood change while holding CBSA white, black and other population shares constant. Doing so avoids including the mechanical effects that rising minority shares have on the education, age, family type and income distributions in these results. These results are reported in Column 9 of Tables 3-6, and are built using the expression

$$f^9_{jt}(i, r, x) = f_{jt}(i|r, x)g_{jt}(x|r)h_{jb}(r).$$

The residual effect (Column 10) is due to changes in racial composition, which typically works against gentrification since the white share of the population has declined over time and downtown neighborhoods have higher minority shares than does the average neighborhood.

Table A1 mathematically specifies construction of each counterfactual distribution and Table A2 reports average shares of target groups across CBSAs overall and within 2 km and 4 km CBD distance rings.

We use the counterfactual distributions $f^c_{jt}(i, r, x)$ and base year distributions $f_{jb}(i, r, x)$ to calculate counterfactual central neighborhood change as follows. Counterfactual population growth g^b_t within 2 or 4 km of CBDs between years b and t reported in Tables 2 and 3 is constructed using the following expression:

$$g^b_t = \frac{1}{J}\sum_j \left(\ln\frac{N_{jt}}{N_{jb}} + \ln\frac{\sum_x\sum_r\sum_{i\subseteq CBD_j} f^c_{jt}(i, r, x)}{\sum_x\sum_r\sum_{i\subseteq CBD_j} f_{jb}(i, r, x)} \right) \tag{3}$$

That is, the central area population growth rate in a CBSA can be expressed as the sum of the CBSA growth rate and the growth rate of the fraction of the population in the central area. The objects reported in Table 2 and 3 are averages across the 120 CBSAs in our sample, as is captured by the outer summation. The reference "no change" results in Column 2 are simply average CBSA population growth rates, calculated as $\frac{1}{J}\sum\limits_j \ln(N_{jt}/N_{jb})$.

Construction of counterfactual white share, college graduate share and median income of neighborhoods within 2 or 4 km of CBDs, appearing in Table 4, is analogous. Exact expressions used to construct these counterfactuals are presented in Appendix B.

Because choices and shares matter multiplicatively for the overall population distribution across tracts, the ordering matters for quantifying the influence of each channel. Table A3 shows results analogous to those in Tables 2 and 3 but imposes the counterfactuals in the reverse order: shares adjustments first and sub-group-specific choice adjustments second. In practice, it shows that the ordering does not materially affect conclusions from this decomposition exercise.

3.2 Counterfactual Results

Before discussing the results of each counterfactual exercise, we summarize the broad picture provided by these decompositions. In the 1980-2000 period, results reflect a pattern of declining central area choice probabilities of all demographic groups except for some young, college educated or high income whites. Departures continued approximately unabated after 2000 among low SES minorities only. In contrast, high-SES whites experienced growing central area choice probabilities and those for low-SES whites stabilized. These changing neighborhood choices account for the bulk of central neighborhood population and demographic change in the 1980-2010 period, with secular demographic change consistently pushing in favor of central area population growth. Central area choice shares of high SES minorities also reversed after 2000 but this group's small population share means their reversal has little effect on aggregates.

Table 2 shows what population growth in 1980-2010 would have been within 2 and 4 km of CBDs under the various counterfactual scenarios laid out in the prior sub-section using race-education joint distributions only. Evidence in Column 1 reiterates the Figure 2 result that populations near CBDs declined until 2000, after which population within 2 km of CBDs grew at about the same rate as overall urban population growth reported in Column 2 while that within 4 km was almost unchanged on average.

Results holding the shares constant in Column 3 are slightly less than the actual changes in Column 1, meaning that shifting demographics pushed toward central area population growth since growing demographic groups were disproportionately located in downtown neighborhoods. Had the race-education distribution not changed from 1980 through 2000, central area population would have declined by 12 percentage points (Column 3) rather than the actual decline of 7 percentage points (Column 1) in the average CBSA. In the 2000-2010 period, central area populations within 2 km of CBDs would have grown by 4 percentage points (Column 3) rather than the 6 percentage points (Column 1) it actually grew. That is, even in the 2000-2010 period, central neighborhood choice probabilities declined in the overall population, with growth in minority shares large enough to counteract these declines and generate small rates of central area population growth. Effects of secular demographic change are roughly the same within 4 km as within 2 km of CBDs.

Table 2 Column 4 shows what would have happened to central area populations had neighborhood choices not changed from base years but demographic shares had. For 1980-2000, it shows 31 percent growth and for 2000-2010, it shows 9 percent growth within 2 km of CBDs. These results reflect the positive effects associated with a rising minority population reinforced by the imposed lack of shifts in neighborhood choices away from central neighborhoods. Comparing the magnitudes of the results in Columns 3 and 4 of Table 2 indicates that changing neighborhood choices has been the key generator of central area population decline in 1980-2000, even as shifting demographics have pushed for central area population growth. In the 2000-2010 period, shifts in neighborhood choices continued to hold central neighborhoods slightly below CBSA growth rates, with demographic change almost making up for this deficit.

Results in Columns 5-8 of Table 2 show the amount of central area population change driven by changes in neighborhood choices by each of the indicated demographic groups. Entries in Columns 5-8 sum to the difference between entries in Columns 2 and 3 (-0.34 for 1980-2000 and -0.03 for 2000-2010 within 2 km of CBDs), or the total impact of changing neighborhood choices holding CBSA demographic composition fixed. These results show that 1980-2000 central area population losses are mostly explained by the flight of less than college educated whites and nonwhites alike, whose effects are similar at -0.14 and -0.18, respectively within 2 km of CBDs. In parentheses is the fraction of each demographic group in the CBSA population. With less than college whites representing the largest shares of CBSA and central area populations, the logic discussed in the context of Equation (2) indicates that the changing choices of less than college nonwhites must be of greater magnitudes. While college educated whites and nonwhites were also choosing to move away from central neighborhoods during 1980-2000, these outflow were much less pronounced and thus contributed little to 1980-2000 central area population declines.

In the 2000-2010 period, minority flight continued and white flight reversed. While less than college educated nonwhites departed central neighborhoods at similar rates as in 1980-2000, whites of all education levels started to return to central neighborhoods. In particular, changing choices of college-educated whites accounted for population growth within 2 km of CBDs of 4 percentage points. Less educated whites were also returning to central areas, but at lower rates than their college-educated counterparts, accounting for 2 percentage points of growth holding shares constant. However, less educated minorities continued to depart central neighborhoods at about the same rate as they had in the prior period, contributing negative 8 points to central area population growth. This evidence of the return of college educated whites to downtown areas is in line with Couture and Handbury's (2016) similar evidence using different census tabulations. We emphasize, however, that it was not enough to counteract the continued departures of less than college educated minorities.

Results in Table 2 Column 9 show how shifts in the composition of the education distribution influenced the central area population share holding racial composition constant. Estimates of -0.04 before 2000 and -0.01 after 2000 indicate declining shares of less educated groups in the population, groups who disproportionately lived in central area neighborhoods in each base year. The results in Column 10 show that the declining white share of the population promoted increases in downtown populations by 10 percentage points in 1980-2000 and 3 percentage points in 2000-2010.

Table 3 reports numbers analogous to those in Table 2, except using joint distributions of age, family type and income with race instead of education. “Target" groups are ages 20-34, singles and couples without children and households in the top 30 percent of the income distribution of our study area. Results in Table 3 are broadly consistent with those in Table 2, with the exception of those using the family type-race joint distribution. Childless households were always prevalent in downtown areas, generating contributions to central area population growth of of 0.10 in 1980-2000 and 0.03 in 2000-2010 holding neigborhood choices fixed, as reported in Column 9. However, childless whites departed central neighborhoods at much higher rates than young and high SES whites during 1980-2000 period. After 2000, however, like young, educated and high income whites, childless whites returned to central neighborhoods, with their changes in neighborhood choices contributing 2 percentage points toward central area population growth. Since most of the 2000-2010 growth in childless households was among minorities, this element of demographic change did not drive much of central area population growth after 2000.

Table 4 reports decompositions of changes in fraction white, fraction college educated and median income of residents within 2 km of CBDs into choice and share based components. It shows why education and income growth before 2000 were leading indicators of racial change in downtown neighborhoods after 2000. While the central mechanisms driving changes in these demographic indicators can mostly be inferred from the population results in Tables 2 and 3, a few observations are of note for the 1980-2000 period. First, secular growth in college fraction accounted for an increase in 0.06 in the fraction of downtown residents with a college education (Panel A, Row 3, Column 9). Second, departures of lower income households from central areas of cities promoted a sizable average increase of 1.8 percentage points in median income of these neighborhoods during this period (Panel A, Row 4, Columns 7 and 8). For the 2000-2010 period, central area median income growth accelarated to 3.8 points, with changes in central neighborhood choices by white high income households contributing 1.8 points to this increase - in addition to persistence in mechanisms that existed before 2000.

4 Understanding Changes in Neighborhood Choices

The prior section presented unified decompositions of the extent to which demographic change in central neighborhoods has been driven by shifts in neighborhood choices by various demographic subgroups. In this section, we interpret this descriptive evidence in the context of a model that ultimately facilitates decompositions of changes in neighborhood aggregate demand into various mechanisms. In particular, this framework allows us to assess the extent to which rising home prices local labor demand shocks and various types of amenities and demand for amenities have driven central neighborhood changes.

4.1 Neighborhood Choice Model

Here we lay out a standard neighborhood choice model that facilitates use of neighborhood choice shares by demographic group along with housing prices to recover information about changes in demand for neighborhoods over time. The procedure makes use of conditional choice probabilities - first formalized in Hotz & Miller (1994) - in a way similar to Bayer et al.'s (2016) dynamic analysis of demand for neighborhood attributes. For clarity of exposition, we begin by thinking about the choice of neighborhood within one CBSA only. Couture & Handbury (2016) show that this is equivalent to considering a nested choice of first CBSA and then neighborhood within the chosen CBSA. Discrete household types are indexed by h and there is a continuum of households of each type.

The indirect utility of household r of type h residing in census tract i at time t is

$$\widetilde{v}^t_{rhi} = v_h(p^t_i, w^t_{hi}, q^t_i) + \varepsilon^t_{rhi} \equiv v^t_{hi} + \varepsilon^t_{rhi}.$$

In this expression, p^t_i is the price of one unit of housing services in tract i, w^t_{hi} is wage net of commuting cost, q^t_i summarizes local amenities and ε^t_{rhi} is an independent and identically distributed (i.i.d.) random utility shock, with a Type I extreme value distribution. q^t_i may be a function of exogenous and endogenous neighborhood attributes including the population composition. w^t_{hi} can depend on human capital characteristics and access to employment locations from tract i. We think of a long-run equilibrium in which moving costs are negligible. This setup delivers the following population shares of household type h in each census tract i, which are observed in the data.

$$\pi^t_{hi} = \frac{\exp(v^t_{hi})}{\sum_{i'} \exp(v^t_{hi'})},$$

suggesting the relationship

$$\ln \pi^t_{hi} = v^t_{hi} - \ln\left(\sum_{i'} \exp(v^t_{hi'})\right). \tag{4}$$

This equation shows that we can use conditional choice probabilities to recover the mean, median or modal utility associated with each tract by each demographic group up to a scale.[4]

We now consider the derivation of estimates of components of indirect utility that capture neighborhood attributes for a reference household type $\overline{h}$ and use it as a basis for recovering such components for other types. The broad goal here is to show how to control for differences in living costs across locations. Impose as a normalization that average modal utility across neighborhoods $\frac{1}{I}\sum_{i'} v^t_{\overline{h}i'} = 1$. This allows for inversion of (4) to an expression relating neighborhood choice probabilities to indirect utility, as in Berry (1994):

$$\ln \pi^t_{\overline{h}i} - \frac{1}{I}\sum_{i'}(\ln \pi^t_{\overline{h}i'}) + 1 = v_{\overline{h}}(p^t_i, w^t_{\overline{h}i}, q^t_i)$$

Fully differentiating yields an expression that tells us that $\ln v_{\overline{h}i}$ equals a weighted average of wages net of commuting costs, home prices and neighborhood attributes relative to those in the average location. This expression assumes utility over goods x, housing H and a local amenity index q, where, $U(x, H, q)$ takes the form $qu(x, H)$, and u is homothetic.

$$\ln \pi^t_{\overline{h}i} - \frac{1}{I}\sum_{i'}\ln(\pi^t_{\overline{h}i'}) = d\ln w^t_{\overline{h}} - \beta_{\overline{h}} d\ln p^t_i + \sigma_{\overline{h}} dq^t_i$$

Here we express utility as relative to the reference location, which has a utility normalized to 1. As in Rosen (1979) and Roback (1982), we see that differences in neighborhood choice probabilities reflect differences in incomes, housing costs and amenity values of locations. We can recover the combination of differences in wages net of commuting costs and local amenities across tracts for the average household type $\overline{h}$ by imposing $d\ln p_i = \ln p_i - \frac{1}{I}\sum_{i'}\ln p_{i'}$.

To recover analogous expressions for household types other than $\overline{h}$, we differentiate indirect utility, holding location constant, to reveal $d\ln v = d\ln w$. Therefore, the reference utility level for households of type h is $1 + \ln w_h - \ln w_{\overline{h}}$, where w_h is the wage

[4] Given the extreme value assumption for the errors, the mean tract utility is $v^t_{hi} + 0.58$ (Euler's constant) with normalization of the scale parameter to 1, the median is $v^t_{hi} - \ln(\ln(2))$ and the mode is v^t_{hi}.

net of commuting cost for type h in the reference (average) location. For generic type h we thus have

$$\ln \pi_{hi}^t + \beta_h d \ln p_i^t = \frac{1}{I} \sum_{i'} \ln(\pi_{hi'}^t) + (\ln w_h^t - \ln w_{\overline{h}}^t) + d \ln w_{hi}^t + \sigma_h dq_i^t \equiv \lambda_{hi}^t. \qquad (5)$$

This formulation takes into account the fact that richer households are more likely to live in higher cost neighborhoods and have lower marginal utilities of income. The result is a greater discount on share differences across locations to reflect the fact that it is less onerous for high-income people to live in high-cost areas than it is for low-income people to live in high-cost areas. It also incorporates type-specific intercepts $\frac{1}{I}\sum_{i'} \ln(\pi_{hi'}^t) + (\ln w_h^t - \ln w_{\overline{h}}^t)$ that we account for empirically using type-CBSA specific fixed effects.

Equation (5) summarizes how to recover the component of differences in neighborhood demands that are driven by differences in wages net of commuting costs and neighborhood amenities. We directly observe π_{hi}^t in the data as $f_{jt}(i|x, r)$ in the context of the counterfactual calculations of the prior section. 0 shares do not match the model well, so we assign tracts with 0 share to the smallest observed positive share for that demographic group for the purpose of calculating shares only. We set valuations of tracts with 0 shares to a missing value. To recover estimates of $d \ln p_i^t$, we take residuals from tract-level regressions of log reported median home price on average home characteristics and CBSA fixed effects in each year.

We aim to construct estimates of β_{hj} (type and CBSA-specific housing expenditure shares), that both reflect potential differences in preferences across groups and that accommodate preferences over housing that may not be homothetic (Albouy, Ehrlich & Liu, 2016). We estimate β_{hj} using data from the 5% public use micro data sample of the 1980 decennial so as to avoid introducing endogenous adjustments to β_{hj} in response to market conditions.[5] To do this, we calculate median type and CBSA specific household level expenditure shares from census micro data and use separate simple regressions of CBSA median share on a CBSA home value index for each group to predict β_{hj}. We choose to calibrate these parameters rather than estimate them because we are dubious about the potential to find clean identifying variation for their estimation. More details about our process for constructing β_{hj} can be found in the data appendix.

[5] A second approach is to instrument for price with attributes of houses and neighborhoods that are located a ways away, as in Bayer, Ferreira & MicMillan (2007), or natural amenities, as in Couture & Handbury (2016). Because natural amenities enter as part of the error term, the second approach does not fit our context well.

Reintroducing the index j for CBSAs, we decompose changes in neighborhood choice probabilities from Equation (5) as follows, where Δ indicates differentials over time and d continues to denote differentials across locations at a point in time:

$$\begin{aligned}\Delta\left(\ln \pi_{hij}\right) = {} & [-\beta_{hj}\Delta(d\ln p_{ij})] + [\Delta\left(d\ln w_{hij}\right)] \\ & +[(dq_{ij}\frac{\Delta\sigma_{hj}}{\Delta\ln w_{hj}} + \sigma_{hj}\frac{\Delta dq_{hj}}{\Delta\ln w_{hj}})\Delta\ln w_{hj}] + [dq_{ij}^{observed}(\Delta\sigma_{hj}|_{\overline{w}_{hj}})] \\ & +[dq_{ij}^{unobserved}(\Delta\sigma_{hj}|_{\overline{w}_{hj}}) + \sigma_{hj}\left(\Delta dq_{ij}|_{\overline{w}_{hj}}\right)] + \rho_{hj}.\end{aligned} \tag{6}$$

Equation (6) shows that changes in neighborhood choice shares reflect shifts in relative cost of living, shifts in relative labor market opportunities, an income effect for local amenities, and shifts in the valuation of existing observed amenities, shifts in valuations of or levels of unobserved amenities and a CBSA-specific trend. Equation (6) implicitly takes into account the fact that demand shifts by high-SES groups push up home prices in certain neighborhoods, thereby dissuading low-SES groups from choosing these neighborhoods, even if their valuations have been rising for other reasons. In Section 4.4, we implement decompositions implied by this expression empirically, with each term in brackets quantified separately.

4.2 Evidence of Relative Demand Shifts for Central Area Neighborhoods

Equation (5) clarifies the existence of equilibrium relationships between decadal changes in log neighborhood choice probabilities adjusted for housing cost changes and factors that influence labor market opportunities and amenities in each tract. We now look to isolate magnitudes of secular relative demand shifts for central neighborhoods and the extent to which these shifts are driven by observed changes in nearby labor market opportunities and consumer amenities. To benchmark the size of these group-specific demand shifts, we first report summary measures of shifts in demand for neighborhoods that incorporate information from all demographic groups simultaneously. We use an index of equally weighted z-scores built using fraction college educated, fraction white and household income as a summary measure of neighborhood demand.

We generalize the logic discussed previously for the Chicago example presented in Figure 3 to each tract in our full sample. In particular, we investigate patterns of changes in central area tracts' demographic composition while accounting for CBSA specific trends in neighborhood inequality and observable natural amenities whose valuations may have changed over time. The following regression equation measures such average differences in central area neighborhood change relative to other neighborhoods.

$$\Delta S_{ijt} = \rho_{jt} + \sum\nolimits_{d=1}^{4} \alpha_{dt} cbddis_{ij}^{d} + \alpha_{1t}^{b} cbddis_{ij}^{1} \Delta \ln Emp_{jt}^{d} + \alpha_{1t}^{s} cbddis_{ij}^{1} \Delta \ln CBDEmp_{jt}^{d}$$
$$+ \sum\nolimits_{d=1}^{4} \beta_{dt} topdis_{ij}^{d} + \sum\nolimits_{m} \delta_{mt} \ln(amendis_{ij}^{m}) + \varepsilon_{ijt} \quad (7)$$

This equation has, ΔS_{ijt}, the change in tract i's SES index (in CBSA j at time t) on the left-hand side as a function of CBSA fixed effects (ρ_{jt}), 4 km CBD distance ring indicators ($cbddis_{ij}^{d}$) with the innermost ring interacted with CBD-oriented ($\Delta \ln CBDEmp_{jt}^{d}$) and CBSA ($\Delta \ln Emp_{jt}^{d}$) labor demand shocks (described later), distance bands to top quartile SES tracts in 1970 ($topdis_{ij}^{d}$) and log distances to various natural amenities ($\ln(amendis_{ij}^{m})$), including coastlines, lakes and rivers. We include controls for natural amenities given evidence in Lee & Lin (2014) that they "anchor" affluent neighborhoods, meaning nearby neighborhoods may be less likely to experience demographic change. The control for distance to top quartile tracts accounts for the possibility that tracts near CBDs gentrified simply because of expansions of nearby high-income neighborhoods (Guerrieri, Hartley, & Hurst, 2013). We estimate coefficients in Equation (7) over each decade 1970-2010 and for the entire 1980-2010 period. We maintain 1970 CBSA population share weights throughout, though we rescale them such that the area within 4 km of each CBSA gets weighted equally across CBSAs.[6]

Table 5 Panel A reports estimates of a_1, a_1^b and a_1^s from Equation (7). α_1 describes how much more or less gentrification occurred in tracts within 4 km of CBDs relative to what was typical among tracts beyond 16 km from the CBD, which is the excluded distance category, quantifying the patterns seen in Figure 2. α_1^b describes how this gap differed by CBSA employment growth $\Delta \ln Emp_{jt}^{d}$. In most periods, we instrument for $\Delta \ln Emp_{jt}^{d}$ using a Bartik (1991) type industry shift-share variable. This instrument is constructed by interacting the 1-digit industrial composition of employment in each CBSA in 1970 with national employment growth rates in each industry to generate a predicted change in employment for each CBSA.[7] The idea is to isolate demand shocks for living in a CBSA that are driven by national trends in industry growth rather than factors that could be correlated with unobservables driving central neighborhood

[6] We assign the region within 4 km of CBDs to have a weight of its average CBSA population share in 1970. We implement this rescaling because our key reported coefficients capture differential impacts on areas within 4 km of CBDs relative to those outside. We wish each CBSA to contribute equally to the identification of these coefficients.

[7] That is, we construct the Bartik instrument for CBSA j that applies to the period $t-10$ to t as: $Bartik_{jt} = \sum_k S_{jk1970} \ln(emp_{kt}^{-j}/emp_{kt-10}^{-j})$, where S_{jk1970} is the fraction of employment in CBSA j that is in industry k in 1970 and emp_{kt}^{-j} is national employment in industry k at time t excluding CBSA j.

change. α_1^s describes how SES growth within 4 km of CBDs differed for CBSAs with larger CBD-oriented labor demand shocks. Here, $\Delta \ln CBDEmp_{jt}^d$ is the change in employment within 4 km of a CBD. $\Delta \ln CBDEmp_{jt}^d$ is instrumented with a CBD-oriented industry shift share variable analogous to the instrument for total CBSA employment growth.[8] So that α_1 can be interpreted as the average demographic change in central area tracts, we standardize $\Delta \ln Emp_{jt}^d$ and $\Delta \ln CBDEmp_{jt}^d$ and their instruments to have means of 0 and standard deviations of 1. Because we do not observe the change in employment within 4 km of CBDs before 1990, we cannot use it as a regressor directly. For this reason (and to maintain consistency across the two Bartik demand shifters) we estimate reduced forms for the 1970-1980, 1980-1990 and 1980-2010 periods instead of instrumental variable (IV) regressions. Therefore, for these periods magnitudes of α_1^b and α_1^s do not accurately capture effects of 1 standard deviation changes in CBSA- and CBD-oriented employment growth, respectively. However, the sign and significance of these coefficients remain informative. Table A4 reports summary statistics about these two types of shocks in each decade, allowing for translation into direct effects of employment changes.

Results in Table 5 Panel A parsimoniously quantify the rebounds experienced by central neighborhoods as visualized in Figure 2, previewing our estimates from the structural model. Our estimate of α_1 in the first row is significantly negative for the 1970s, becomes near 0 for the 1980s and 1990s, and strengthens further in the 2000-2010 period to show that on average central areas experienced 0.14 standard deviations more positive demographic change than the typical suburban neighborhood. Over the longer 1980-2010 period, central areas experienced 0.17 standard deviations more positive demographic change relative to suburban neighborhoods. Because the interactions terms are normalized to be mean 0, the interpretation of this first row of coefficients is as an average across CBSAs.

The second and third rows present estimates of α_1^b and α_1^s, respectively. One consistent finding is that central neighborhoods of CBSAs with more robust central area employment growth experienced relatively more gentrification (seen in the positive α_1^s coefficients), even in the 1970s. However, this phenomenon was strongest in the 2000-2010 period when 1 standard deviation greater downtown employment growth

[8] For CBSA j, denote the fraction of employment near the CBD in industry k in 1990 as f_{jk}^{emp}. We think of f_{jk}^{emp} as being driven by the interaction of fundamental attributes of the production process like the importance of agglomeration spillovers to total factor productivity (TFP). Therefore, we predict the change in the fraction of employment near the CBD to be $Spatbartik_{jt} = \sum_k f_{jk}^{emp} \ln(emp_{kt}^{-j}/emp_{kt-10}^{-j})$, where emp_{kt}^{-j} denotes national employment in industry k and year t excluding CBSA j.

generated a 0.12 standard deviation relative increase in central area SES index. (These coefficients only have clean interpretations for the 90s and 00s when we can estimate them by IV.) The effects of CBSA employment growth on downtown neighborhood change depend a lot more on the time period and better track average trends. In the 1970s, central areas of CBSAs with more robust exogenous employment growth deteriorated more than was typical, whereas by 2000-2010 the reverse was true, though our estimate is not statistically significant. That is, broader forces buffeting central area neighborhoods appear to be reinforced by trends in aggregate CBSA labor demand shocks. Similar patterns are found separately within each tercile of the 1970 SES index distribution. That is, these results are not only being driven by low-SES central neighborhoods.

Evidence from Chicago in Figure 3 reveals that neighborhoods experienced mean reversion in their SES index. This mean reversion is pervasive across CBSAs, and it is relevant to our setting because central area tracts disproportionately appear in the bottom half of the SES index distribution. We account for mean reversion in two ways. First, we include an additional control for S_{ijt-10} in Equation (7) and consolidate S_{ijt-10} onto the right-hand side of the regression equation. This yields an AR(1) specification with CBSA fixed effects fully interacted with the lagged SES index. This specification generates regression lines for each CBSA*decade combination analogous to those in Figure 3 for Chicago.

$$\begin{aligned} S_{ijt} = \rho'_{jt} + \mu'_{jt} S_{ijt-10} + \sum\nolimits_{d=1}^{4} \alpha'_{dt} cbddis^{d}_{ij} + \alpha^{b'}_{1t} cbddis^{1}_{ij} \Delta \ln Emp^{d}_{jt} + \alpha^{s'}_{1t} cbddis^{1}_{ij} \Delta \ln CBDEmp^{d}_{jt} \\ + \sum\nolimits_{d=1}^{4} \beta'_{dt} topdis^{d}_{ij} + \sum\nolimits_{m} \delta'_{mt} \ln(amendis^{m}_{ij}) + \varepsilon'_{ijt} \qquad (8) \end{aligned}$$

These regressions feature the same remaining set of regressors as in (7). Table 5, Panel B reports estimates of coefficients in Equation (8).

While this empirical approach addresses mean reversion, it is well known that in short panels OLS estimates of μ_{jt} may be biased downward. Such "Hurwicz bias" will influence coefficients of interest α_1, α^b_1 and α^s_1 if the lagged SES index is correlated with CBD distance, which is likely as CBD areas are more likely to be poor - the whole justification for exploring this specification from the start. To deal with this bias, we implement a standard Arellano-Bond (1991) type correction. Beginning with (8), impose that $\mu_{jt} = \mu_{jt-1}$ and, without loss of generality, add a tract fixed effect.

First-differencing yields the following equation:

$$\Delta S_{ijt} = \rho''_{jt} + \mu''_{jt}\Delta S_{ijt-1} + \sum{}_{d=1}^{4}\alpha''_{dt}cbddis_{ij}^{d} + \alpha_{1t}^{b\prime\prime}cbddis_{ij}^{1}\Delta\ln Emp_{jt}^{d} + \alpha_{1t}^{s\prime\prime}cbddis_{ij}^{1}\Delta\ln CBDEmp_{jt}^{d} + \sum{}_{d=1}^{4}\beta''_{dt}topdis_{ij}^{d} + \sum{}_{m}\delta''_{mt}\ln(amendis_{ij}^{m}) + \varepsilon''_{ijt} \quad (9)$$

As in the standard Arellano-Bond (1991) procedure, we instrument for ΔS_{ijt-1} with S_{ijt-2}. The identifying assumption is thus that the lagged SES index is not correlated with unobservables driving innovations in a tract's SES index after accounting for mean reversion, CBD distance and distance to amenities. In practice, this means we have J instruments, one for each CBSA interacted with ΔS_{ijt-1}. Results from this specification are reported in Table 5, Panel C, with 1970-1980 omitted because data from 1960 are not available to form instruments for these estimates.

The results in Table 5 Panels B and C are quite similar to those in Panel A. Whichever assumption we impose about the underlying data-generating process, the three main facts persist. First, there is a clear statistically meaningful demographic rebound of central neighborhoods in the 2000-2010 period. Second, central area employment growth bolstered central neighborhood demographic change, especially in the 1970-1980 and 2000-2010 periods. Third, CBSA employment growth bolstered central neighborhoods only in the 2000-2010 period, when they were changing for other reasons. The results in Table 5 Panels B and C demonstrate that the reversal of fortune experienced by many central neighborhoods after 1980 is not purely an artifact of mean reversion.[9]

Overall, the evidence in Table 5, as well as facts about central area employment growth, indicates that the bulk of 2000-2010 downtown gentrification could not have been driven by shifts in the spatial structure of labor demand. However, CBD-oriented positive labor demand shocks reinforced the downtown gentrification that occurred in many cities primarily for other reasons. With 2000-2010 CBD area employment growth averaging -1 percent across CBSAs, downtown neighborhood growth must have come about for other reasons in most CBSAs, with improvements in the relative amenity values of downtown neighborhoods the most logical mechanism.[10] [11]

[9] As an alternative for examining the impact of mean reversion, we generated results as in Table 5 Panel A for tracts in terciles of the 1970 SES distribution. We get similar results for the top and bottom terciles, further evidence that mean reversion is not driving the results.

[10] Regression results analogous to those in Table 6 using an index of tract housing value growth rates as the dependent variable give similar results. These results appear in Table A5.

[11] Edlund, Machado, & Sviatchi (2015) find that 26 large CBSAs with stronger skilled labor Bartik shocks experienced more rapid decadal central home price growth and demographic change in central areas than other areas of the city. These patterns are replicated in our data as well if census tracts are

4.3 Using the Model

Figure 4 gives a sense of how tract valuations λ_{hi}^t in (5) have changed since 1980 as functions of CBD distance for four demographic groups. It shows the average change across CBSAs in calibrated versions of λ_{hi}^t for 0.5 km CBD distance rings. $\Delta\widehat{\lambda}_{hij}^t$ are constructed using tract choice shares, housing expenditure shares and home price indexes. Figure 4 shows that college whites and blacks and high school dropout whites and blacks all experienced rising valuations of neighborhoods within 2 km of CBDs after 2000, though the estimates for the blacks are much noisier given their small population shares. However, comparing results in Panel A to those in other panels reveals that college whites have valuations that increase the most over the broadest distance range, out to about 3 km from CBDs. Next, we evaluate the drivers of these changes and their implications for central area population and demographic changes.

We investigate the extent to which CBSA-level and localized labor demand shocks have driven changes in neighborhood valuations using regression equations similar to Equation (8) for each demographic group separately. We think of CBD-oriented labor demand shocks as influencing $d\ln w_{hi}^t$ and CBSA-level labor demand shocks as potentially changing groups' demands for central area amenities through an income effect, as laid out in (6). We report IV regression results from estimating the following equation for the 1990-2000 and 2000-2010 periods, since we only observe the change in employment near CBDs starting in 1990. For other time periods, we report the reduced form. The specification is as follows:

$$\begin{aligned}\Delta\widehat{\lambda}_{hij}^t = \rho_{hjt} + \sum\nolimits_{d=1}^4 a_{hdt} cbddis_{ij}^d + a_{h1t}^b cbddis_{ij}^1 \Delta \ln Emp_{jt} + a_{h1t}^s cbddis_{ij}^1 \Delta \ln CBDEmp_{jt} \\ + \sum\nolimits_{d=1}^4 b_{hdt} topdis_{ij}^d + \sum\nolimits_m c_{hmt} \ln(amendis_{ij}^m) + e_{hijt}. \qquad (10)\end{aligned}$$

This estimation equation is the empirical analog to a differenced version of Equation (5). ρ_{hjt} accounts for the intercept $\frac{1}{I}\sum_{i'} \ln(\pi_{hi'}^t) + (\ln w_h^t - \ln w_{\overline{h}}^t)$, and the remaining terms allow us to measure variation in tract-level labor market opportunities and local amenities relative to those in the average location. Here, we no longer impose that $\Delta \ln Emp_{jt}$ and $\Delta \ln CBDEmp_{jt}$ have means of 0, though we maintain standard deviations of 1. As a result, a_{hdt} represents average demand shifts for central neighborhoods that occur for unobserved reasons only. Because we use parameter estimates to perform a unified accounting of mechanisms driving central neighborhood change, we need a consistent sample over time for each demographic group. To achieve this,

equally weighted.

we only use tracts with nonzero choice shares in all years 1980-2010 for estimation of (10). Observations are weighted by the inverse of the number of CBSA tracts in the sample, and regions within 4 km of the CBD is given equal weight in each CBSA, so that each CBSA and its central area is weighted equally. This way, key parameters a_{hdt}, a^b_{h1t} and a^s_{h1t} represent average influences across all CBSAs in our sample. Table A4 reports descriptive statistics about CBD-area and CBSA employment changes and their instruments.

There are two potential concerns with using Equation (10) to infer reasons for changes in neighborhood valuations. First is the issue of whether we have accurately measured housing costs. To get around this, instead of Equation (10) one could estimate a unified equation for all household types simultaneously with type by tract fixed effects. Because the housing cost is common across types, the tract fixed effect would control for these costs if the housing expenditure share were the same for all types. The costs of this approach are that the absolute change in tract valuation is lost to a normalization, meaning that one can only recover relative changes in tract valuations across demographic groups, and that housing expenditure shares empirically differ across groups. Our experimentation with such unified regression specifications yield very similar conclusions about relative changes in central area tract valuations across demographic groups to the results reported in Table 6.

In addition to providing evidence about mechanisms driving shifts in central area valuations, estimates of coefficients in (10) are used to carry out unified decompositions of mechanisms driving numbers reported in Table 2 Columns 5-8. Such decompositions require us to be able to calculate neighborhood choice probabilities for each demographic group for each tract in each sample year under various counterfactual scenarios. Because the analysis is carried out in logs, there is a difficult issue of what to do about neighborhoods with 0 choice shares. Our solution is to exclude any tract from the estimation sample if it had a 0 neighborhood choice share in any year 1980-2010, applying this rule separately by demographic group. While this restriction means we do not use potentially useful information about increasing group demand for tracts going from 0 to positive choices shares and vice versa, it is needed to use these results to carry decompositions that apply to a consistent geography over time. As a robustness check, we estimate versions using a data set in which all tracts within 2 km CBD bands are combined into a single observation per group per CBSA. Results using this aggregate data set are very similar to the results presented in Table 6.

The results in Table 6 show that each group's demand for central area residency is estimated to respond positively to central area employment shocks, particularly in the 2000-2010 period. However, all group's demands respond negatively to CBSA level employment shocks (conditional on CBD-oriented employment shocks), except college educated whites. Conditional on nearby labor market opportunities, rising incomes drove suburbanization for all but college educated whites. This result is in contrast to evidence from prior decades. While responses to CBSA employment shocks are mixed across groups in the 1990s (and not statistically significant in any case), in the 1980s, each group's central neighborhood demand is estimated to respond negatively to CBSA employment shocks. This is evidence that central neighborhoods were mostly inferior goods for all groups prior to 1990, but became normal goods for college educated whites after 2000. Coefficients in the top row of each panel of Table 6 indicate additional changes in central area demands that are due to unobserved factors. These estimates are consistently negative across groups in the 1980s and are greater for college graduates than others in the 2000-2010 periods.[12]

The results for whites and blacks who completed high school but not college (not reported in Table 6) are in between the college graduate and high school dropout results for each race. Conditional on educational attainment, the results for the "other" demographic group are between those for whites and blacks, though somewhat more similar to those for whites. We repeat the same exercise as in Table 6 using income deciles instead of education groups, with results quite similar to those in Table 6. The background changes in central neighborhood valuations improved more for the high deciles than for the low deciles, but only turned significantly positive for high-income whites, not blacks.

4.4 Decompositions of Shifts in Neighborhood Choices

With education-race specific estimates from (10) in hand, we combine insights from the model with estimates similar to those in Table 6 for each education-race group to generate unified decompositions of the relative importance of various mechanisms driving shifts in downtown neighborhood choices. This exercise decomposes the contributions of shifts in neighborhood choices by four education-race groups to central neighborhood population change, reported in Table 2 Columns 5-8, into 6 components: home price changes, CBD-oriented labor demand shocks, CBSA labor demand shocks, central area fixed effects, and a residual. Combining Equations (5) and (10), we have the following

[12] Reported standard errors are overstated due to clustering at the CBSA level. Future versions will report spatially corrected standard errors instead.

decomposition of shifts in the log share of group h choosing to live in census tract i

$$\begin{aligned}\Delta \ln \pi^t_{hij} = &\left[-\beta_{hj}\Delta d \ln p^t_{ij}\right] + \left[\alpha^s_{h1t} cbddis^1_{ij} \Delta \ln CBDEmp_{jt}\right] + \left[\alpha^b_{h1t} cbddis^1_{ij} \Delta \ln Emp_{jt}\right] \\ &+ \left[\sum{}^4_{d=1} \beta_{hdt} topdis^d_{ij} + \sum{}_m \delta_{hmt} \ln(amendis^m_{ij})\right] \\ &+ \left[\sum{}^4_{d=1} \alpha_{hdt} cbddis^d_{ij}\right] + [\varepsilon_{hijt}] + \rho^t_{hj} \qquad (11)\end{aligned}$$

Parameters α^s_{h1t}, α^b_{h1t}, β_{hdt}, δ_{hmt}, ρ^t_{hj} and α_{hdt} are estimated in the context of (10), as reported for four of the 12 race-education groups in Table 6, while β_{hj} is calibrated as described above.

Each bracketed term in (11) has a substantive interpretation in the context of Equation (6) from the model, where each parameter should be interpreted as an average for central areas across CBSAs. $\left[-\beta_h \Delta d \ln p^t_{ij}\right]$ captures effects of housing cost changes $-\beta_{hj}\left[\Delta\left(d \ln p_{ij}\right)\right]$. $\left[\alpha^s_{h1t} cbddis^1_{ij} \Delta \ln CBDEmp_{jt}\right]$ corresponds to relative shifts in central area employment opportunities $\Delta\left(d \ln w_{hij}\right)$. $\left[\alpha^b_{h1t} cbddis^1_{ij} \Delta \ln Emp_{jt}\right]$ corresponds to income growth's differential effect on downtown neighborhood demand through its impact on preferences for amenities and amenities themselves $(dq_{ij} \frac{\Delta \sigma_{hj}}{\Delta w_{hj}} + \sigma_{hj} \frac{\Delta dq_{hj}}{\Delta w_{hj}})\Delta w_{hj}$. $\left[\sum{}^4_{d=1} \beta_{hdt} topdis^d_{ij} + \sum{}_m \delta_{hmt} \ln(amendis^m_{ij})\right]$ corresponds to changes in the valuation of observed neighborhood amenities holding incomes constant $dq^{observed}_{ij}(\Delta \sigma_{hj}|_{\overline{w}_{hj}})$. The dummy variable for neighborhoods within 4 km of a CBD $\alpha_{hdt} cbddis^1_{ij}$ captures average changes in the valuation of central area unobserved neighborhood amenities relative to those beyond 12 km from CBDs holding incomes constant, corresponding to $dq^{unobserved}_{ij}(\Delta \sigma_{hj}|_{\overline{w}_{hj}})$ and amenity changes in central areas holding incomes constant $\sigma_{hj}\left(\Delta dq_{ij}|_{\overline{w}_{hj}}\right)$. The error term ε_{hijt} captures idiosyncratic reasons why neighborhood choice shares changed, which are mean 0 within CBSAs and within central areas. This error component will matter because of the nonlinear transformation needed to recover levels of neighborhood choice probabilities π^t_{hij}. Finally, there is a CBSA fixed effect that will be ignored, as we must rescale all CBSA neighborhood choice probabilities to sum to 1 in any case.

To carry out decompositions, we construct a series of counterfactual census tract choice shares for each education-race group in 2000 and 2010, taking 1980 and 2000 neighborhood choice shares as given. To build counterfactual group specific year 2000 neighborhood choice shares $\pi^{2000,1}_{hij}, \pi^{2000,2}_{hij}$, etc. we apply regression results from the 1980s and the 1990s sequentially. Because we do not observe central area employment growth between 1980 and 1990, yet reduced form coefficients on Bartik variables are small and not significantly different from 0 for all groups as seen in Table 6, we set a^s_{h1t} in (10) to 0 and estimate (10) as an IV regression for the 1980s. For following decades, we estimate (10) by IV as written for each group.

Counterfactual neighborhood choice shares incorporate each component described above one by one. For example, counterfactual 2010 neighborhood choice shares that only incorporate 2000-2010 housing price changes are calculated as

$$\pi^{2010,1}_{hij} = s^{2010,1}_{hj} \pi^{2000}_{hij} e^{-\beta_h \Delta d \ln p^t_{ij}}. \tag{12}$$

In this expression, $s^{2010,1}_{hj}$ is a group-CBSA specific scale factor that is set to ensure that group-specific neighborhood choice shares sum to 1 within each CBSA.[13] Counterfactual 2010 neighborhood choice shares that incorporate additional mechanisms include additional components of (11) in the exponential component of (12). Because $E_{i|central}\left[\exp(\varepsilon_{hijt})\right]$ is not mean 0 and because of the Jacobian transformation associated with exponentiation of a log normal random variable, there can be positive or negative unexplained components of the population growth. For each set of counterfactual choices $\pi^{y,c}_{hij}$, we form a data set and recalculate Columns 5-8 of Table 2 using each of these counterfactual data sets. Contributions of each component discussed above to each entry in these columns of Table 2 are reported in Table 7.

Table 7 presents the components of population growth within 2 km (left side) or 4 km (right side) of CBDs. Note that entries in the "Total" row do not exactly match numbers in Table 2 columns 5-8 because the sample used to estimate components in Table 7 is slightly more restrictive than the full set of tracts used to construct numbers in Table 2 and excludes 3 CBSAs for which we have no information on observed amenities. Each component listed in the table corresponds to a term in brackets in Equation (11) in the same order. The entries are calculated as follows. We estimate separate regressions using Equation (10) like those used to create Table 7 for each decade and narrowly defined education-race group. Then, components are cumulatively added to log neighborhood choices shares from the base year following Equation (11), exponentiated and normalized to sum to 1 for each demographic group across census tracts in each CBSA. Table 7 expresses results as marginal contributions of each listed mechanism to the component of central area population growth that is due to shifts in the indicated demographic group's change in neighborhood choices. Therefore, each column within the left block of each panel sums to numbers in Table 2, Columns 5-8.

[13] The size of this normalization has no effect on results, as the ultimate statistics we calculate are for each central area relative to other areas of the same CBSA.

Results in Table 7 Panel A indicate that the largest force driving 1980-2000 central neighborhood population declines were a combination of reductions in the quality of unobserved amenities and changes in the marginal utilities of existing amenities, except among less than college whites. It is particularly striking that this force drove -0.13 of the -0.20 for less than college nonwhites within 2 km of CBDs. The other important force for those with less than a college education only is the income effect on demand for central amenities. For whites, this component accounts for -0.10 of -0.14, with unobserved amenities and changes in employment opportunities next in importance at about -0.025 each. Forces examined other than unobserved amenities had little impacts on changing 1980-2000 neighborhood choices of the college educated.

Results in Panel B indicate the continued importance after 2000 of changes in the valuation of and/or level of unobserved amenities for driving changes in central neighborhood choice probabilities for all groups. However, only less than college educated minorities continued to value these changes negatively. At 0.013 out of a total impact of 0.03, our results indicate that unobserved amenities is the most important factor driving the return of white college graduates to central areas of cities, echoing Couture & Handbury's (2016) evidence. Shifts in the valuations of unobserved amenities by less educated whites also mattered a lot for population turnarounds in central neighborhoods. The impact of 0.021 for this group counteracted negative impacts of declining valuations of observed amenities and a negative income effect for existing central area amenities.

Evidence for regions within 4 km of CBDs, reported in the right block of Table 7, is very much in line with that for within 2 km of CBDs.

5 Conclusions

Neighborhoods near central business districts of U.S. metropolitan areas have experienced remarkable rebounds in population and their residents' socioeconomic status since 2000. Decompositions reveal that this turnaround in population has primarily been driven by the return of college-graduate and high-income whites to these neighborhoods, coupled with a halt in the outflows of other white demographic groups. At the same time, the departures of minorities without college degrees continued unabated.

Estimates from our neighborhood choice model indicate that better nearby labor market opportunities draw in residents, but conditional on such opportunities higher incomes only drew in more college graduates to central neighborhoods in the 2000-2010 period. However, we find that most groups except less than college educated minorities experienced growth in their valuations of central area unobserved amenities in 2000-2010 after declines in the 1980-2000 period. Our quantitative decompositions of the mechanisms driving central area population change reveals that 1980-2000 reductions in the quality of central neighborhood amenities was the most important driver of these areas' population declines except amongst low SES whites, for whom income effects on existing local amenities were most important. For the 2000-2010 period, income effects were the most important driver of continued departures of low SES minorities, while the positive impacts of unobserved amenities was most important for other groups. While all groups value improved downtown labor market opportunities, the average CBSA experienced declining downtown employment in the 1990s and essentially no change in downtown employment in the 2000-2010 period. As a result, shifts in central area labor market opportunities had a miminal impact on central area population changes since 2000, though the stabilization of downtown employment declines represents a force that promoted stabilization after 1980-2000 downtown population declines.

The gentrification of cities' central neighborhoods inverts the decentralization of high-income whites that had been occurring for decades. This represents a fundamental change in the demographic structure of cities, for which this paper provides only a starting point from which to build a deeper understanding. This phenomenon may be the beginning of an urban rebirth with many broader consequences for the economy. It may also exacerbate the rise in real income inequality that has occurred over recent decades, as it is a mechanism through which the cost of living may be rising for the poor. A general equilibrium framework which incorporates housing supply would be required to recover information about associated welfare consequences. Developing such a framework which could be used to evaluate the welfare consequences of gentrification for poor incumbents, seems like a particularly fruitful area for future research.

A Data Appendix

A large portion of the data used in our analysis come from tract-level tabulations from the Decennial Census of Population and Housing for the years 1970, 1980, 1990, and 2000, and from the American Community Survey (ACS) for the years 2008-2012. We use census tract boundaries from the 2000 census. We begin with the normalized data provided in Geolytics' 1970-2000 Neighborhood Change Database (NCDB) which provides a subset of the tract-level tabulation variables available from the 1970, 1980, 1990, and 2000 censuses normalized to year 2000 tract boundaries. We augment this data with other tract-level tabulations from these censuses that are not available in the

NCDB and tract-level estimates from the 2008-2012 ACS. In these cases, we perform normalizations to 2000 tract boundaries using the appropriate census tract relationship files provided by the U.S. Census Bureau.[14]

A.1 Tract-level Sample

Our sample includes all of the 2008 definition Core Based Statistical Areas (CBSAs) that had a population of at least 250,000 in the area that was tracted in 1970 except Honolulu.[15] Our sample consists of 120 CBSAs and 39,087 year 2000 census tracts.[16] The CBSAs in the sample can be seen in Figure 1.

A.1.1 1970, 1990, and 2000 Tract Data

We take these directly from the Neighborhood Change Database (NCDB) STF3A tabulations.

A.1.2 1980 Tract Data

We read in these data from the summary tape file 4 files. This allows us to incorporate household income distributions by race and age by race into the data set. It also facilitates imposing various appropriate adjustments for suppression that are not handled well in the NCDB.

Suppression results in undercounting of whites and blacks in various tables. To handle this, we use tract-level full population or household counts of whites, blacks and others to form inflation factors. We calculate inflation factors that scale up the total number of people in each age, education, family type or income bin in the STF4A data to equal the total reported in the NCDB data.

In particular, in the case of age, when the 1980 STF4A tract tabulations by race and age do not sum to the total population, we implement the following algorithm:

1. Inflate the total in each age bin so that the total of the age bins sums to the total population in the NCDB data.

[14] See https://www.census.gov/geo/maps-data/data/relationship.html

[15] Since we are using year 2000 tract boundaries, we limit our sample slightly further by using only tracts for which 100% of the 2000 definition tract was tracted in 1970.

[16] For CBSAs that are split into Metropolitan Divisions, we treat each Division as a separate entity except in the following 4 cases in which we combine Metropolitan Divisions. The 4 cases are as follows: 1) Bethesda-Rockville-Frederick, MD, is combined with Washington-Arlington-Alexandria, DC-VA-MD-WV; 2) Cambridge-Newton-Framingham, MA, and Peabody, MA Metropolitan Divisions are combined with Boston-Quincy, MA; 3) Nassau-Suffolk, NY, is combine with New York-White Plains-Wayne, NY-NJ; and 4) Warren-Troy-Farmington Hills, MI, is combined with Detroit-Livonia-Dearborn, MI.

2. Calculate other race in each age bin by taking the total population in each age bin and subtract the white and black population of that age bin from the STF4A.

3. Calculate the number of whites and blacks that are missing in the STF4A data by summing across the age bins for white and for black and subtracting the totals from the NCDB totals.

4. Calculate the number of people missing from each age bin by subtracting the STF4A total (that uses the recalculated other category) from the NCDB total.

5. Inflate the number of others in each age bin by the ratio of the NCDB other total to the STF4A other total.

6. Calculate the residual number of blacks and whites missing from each age bin by subtracting the inflated other from the inflated total for the age bin.

7. Reassign the residual number of blacks and whites missing from each age bin to either the white or black count in proportion to the share of the total missing that white and black make up as calculated in 3.

We perform the same process for education and family type in 1980.

A.1.3 2010 Census and ACS

We use the 2010 census summary tape file 1 for information about age and household structure by race. Because of the lack of a census long form in 2010, we are forced to use the ACS to measure joint distributions of race by education and race by income.

A.2 Procedure for Allocating Income To Percentile Bins

The counterfactual analysis uses 10 household income deciles, with dollar cutoffs calculated using census micro data for the CBSAs in our sample. In each year, the census tract data reports the number of households by race in each of up to 20 income bins bounded by fixed dollar cutoffs. To re-allocate into percentile bins, we assume a uniform distribution within each dollar value bin except the top one. For the top one, we use a Pareto distribution with parameters estimated separately for each year using census micro data.

A.3 Central Business District Definitions

For each of our 120 CBSAs, we define the Central Business District (CBD) of the CBSA as that of the most populous Census place within the CBSA based on the year 2000 population. We make two exceptions to this rule based on our knowledge of the cities. For the Santa Barbara-Santa Maria-Goleta, CA Metropolitan Statistical Area we use the Santa Barbara CBD rather than the Santa Maria CBD even though Santa Maria was more populous in 2000 than Santa Barbara. For the Virginia Beach-Norfolk-Newport News, VA-NC Metropolitan Statistical Area we use the Norfolk CBD rather than the Virginia Beach CBD. For 113 of the our 120 CBSAs we were able to determine the CBD of the most populous city from the 1982 Census of Retail Trade. We use the latitude and longitude of the centroid of the tract or tracts specified as CBD tracts. For the remaining 7 CBSAs, we used the latitude and longitude as designated by the mapping software maker ESRI.[17]

A.4 Bartik Instrument Construction

We construct two Bartik instruments from several data sources. We label these instruments "Employment Bartik" and "Spatial Employment Bartik." The "Employment Bartik" attempts to predict CBSA-level employment growth for each of the 4 decades using initial year employment shares and decadal employment growth (implemented as changes in log employment levels) with 10 broad industry categories that can be consistently constructed from 1970 through 2010 using the county-level U.S. Census and ACS tabulations. The 10 industry categories are: 1) Agriculture, forestry, fisheries, and mining; 2) Construction; 3) Manufacturing; 4) Wholesale trade; 5) Retail trade; 6) Transportation, communication, other public utilities, and information; 7) Finance, insurance, and real estate; 8) Services; 9) Public administration; and 10) Military. We refer to these as 1-digit industry categories.[18] This measure uses the exact geographical boundaries included in each of our CBSA definitions over the entire time period.

The aim of the "Spatial Employment Bartik" is to predict which CBSAs might be particularly affected near the CBD by national industry growth. To construct this index, we calculate the share of employment located within 4 km of the CBD made up by each industry for each CBSA, using the year 2000 Census Transportation Planning Package. We take these shares and interact them with the national industry growth rate of that industry to form a spatial or CBD-focused Bartik instrument. Ideally, we would calculate the shares in each initial year, 1970, 1980, 1990, and 2000. However, the data are only available starting in 1990. Therefore, we use the 1990 1-digit industry distribution as the base.

[17] These 7 cities are Duluth, MN, Edison, NJ, Indianapolis, IN, Jacksonville, FL, Nashville, TN, and York, PA. Manual inspection of these 7 cities revealed CBD placement where we would expect it. Also, for the 113 cities where we have both Census of Retail Trade and Esri CBD definitions, the points line up closely.

[18] In practice, we do this once for each CBSA excluding that CBSA to calculate a national-level change that is not influenced by that particular CBSA.

A.5 Construction of Housing Expenditure Shares β_{hj}

To construct estimates of β_{hj} (type and CBSA-specific housing expenditure shares) we begin with a sample of renters and owner-occupier households with a mortgage that moved in the 5 years leading up to 1980 and are not living in group quarters. This group experiences housing costs that are closest to 1980 market conditions. We include all mortgage payments, rent, utilities and insurance in housing costs. We trim the 1st and 99th percentiles of housing cost and the 1st and 99th percentiles of household income and take their ratio to calculate the housing expenditure share for each household. Next, we calculate the median expenditure share for each race - educational attainment - CBSA cell. Since some of the cell sizes are quite small we use the predicted values from a linear regressions of housing expenditure shares on a CBSA home value index within each race - educational attainment combination. Resulting housing expenditure shares range from 0.20 to 0.37.

B Construction of Counterfactuals in Table 4

We calculate changes in central areas' white and college-graduate shares using the following expressions respectively. The associated results appear in rows 1-3 of each panel of Table 4.

$$\frac{1}{J}\sum_j \left(\frac{\sum_x \sum_{i\subseteq CBD_j} f^c_{jt}(i,r=w,x)}{\sum_x \sum_r \sum_{i\subseteq CBD_j} f^c_{jt}(i,r,x)} - \frac{\sum_x \sum_{i\subseteq CBD_j} f_{jb}(i,r=w,x)}{\sum_x \sum_r \sum_{i\subseteq CBD_j} f_{jb}(i,r,x)} \right) \tag{13}$$

$$\frac{1}{J}\sum_j \left(\frac{\sum_r \sum_{i\subseteq CBD_j} f^c_{jt}(i,r,x=\mathrm{col})}{\sum_x \sum_r \sum_{i\subseteq CBD_j} f^c_{jt}(i,r,x)} - \frac{\sum_r \sum_{i\subseteq CBD_j} f_{jb}(i,r,x=\mathrm{col})}{\sum_x \sum_r \sum_{i\subseteq CBD_j} f_{jb}(i,r,x)} \right). \tag{14}$$

In these expressions, x indexes education group or income decile as indicated in the row label of Table 4. $i \subseteq CBD_j$ indicates a summation only over tracts within 2 or 4 km of CBSA j's CBD. The reference change for both outcomes is 0 (Column 2 of Table 4), since there is no scale component.

Remaining rows in Table 4 report counterfactual changes in central area median household income. We use median rather than mean income in order to be more robust in avoiding misallocating households into incorrect income deciles.[19] To see how these medians are constructed, begin with the following expression for the cumulative

[19] Because cutoffs associated with each decile do not match the dollar cutoffs in the tract data, we assume uniform distributions within census data dollar bands for allocation purposes. The Data Appendix details our procedure for allocating households to income deciles.

distribution function of CBSA j's central area households across income deciles $x \subseteq \{1, 2, ..., 10\}$.

$$G_{jt}^c(X) = \frac{\sum_{x \leq X} \left[\sum_r \sum_{i \subseteq CBD_j} f_{jt}^c(i, r, x) \right]}{\sum_x \sum_r \sum_{i \subseteq CBD_j} f_{jt}^c(i, r, x)}.$$

The income deciles are defined for the full national study area, but here we only focus on the cumulative distribution function (cdf) for central neighborhoods under counterfactual c. Using these distributions over deciles x, we identify the deciles D_{jt}^c that contain 0.5. We assign the median percentile assuming a uniform distribution of household income within D_{jt}^c. For example, if $G_{jt}^c(2) = 0.45$ and $G_{jt}^c(3) = 0.55$, $D_{jt}^c = 3$. In this case, we would assign the median household income M_{jt}^c in CBSA j at time t under counterfactual c to be 25, representing the 25^{th} percentile of the full study area's household income distribution. Then, the statistics reported in Table 4 are

$$\frac{1}{J} \sum_j \left(M_{jt}^c - M_{jb} \right). \tag{15}$$

As a result, positive numbers in Table 4 mean that the counterfactual in question pushed central area median incomes up by the indicated number of percentile points out of the national urban household income distribution.

References

Ahlfeldt, Gabriel, Stephen Redding, Daniel Sturm, Nikolaus Wolf (2015) "The Economics of Density: Evidence from the Berlin Wall," Econometrica 83(6): 2127-2189.

Altonji, Joseph, Ulrich Doraszelski, and Lewis Segal (2000) “Black/White Differences in Wealth,” Economic Perspectives, Vol. 24(1): 38–50.

Arellano, Manuel, and Stephen Bond (1991) "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations." The review of economic studies 58.2, 277-297.

Bartik, Timothy (1991) Who Benefits from State and Local Economic Development Policies? Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Baum-Snow, Nathaniel (2007) “Did Highways Cause Suburbanization?” Quarterly Journal of Economics, 122(2), 775-805.

Baum-Snow (2017) “Urban Transport Expansions, Employment Decentralization, and the Spatial Scope of Agglomeration Economies,” manuscript.

Bayer, Patrick, Fernando Ferreira & Robert McMillan (2007) “A Unified Framework for Measuring Preferences for Schools and Neighborhoods” Journal of Political

Economy 115(4): 588-638.

Bayer, Patrick, Robert McMillan, Alvin Murphy, & Christopher Timmins (2016) “A Dynamic Model of Demand for Houses and Neighborhoods.” Econometrica, 84:3.

Berry, Steven (1994) “Estimating Discrete-Choice Models of Product Differentiation,” The RAND Journal of Economics, 25 (2), 242–262

Boustan, L.P. (2010) "Was Postwar Suburbanization "White Flight"? Evidence from the Black Migration." Quarterly Journal of Economics 125.1: 417–43.

Brueckner, J. K., & Rosenthal, S. S. (2009) "Gentrification and neighborhood housing cycles: will America’s future downtowns be rich?" The Review of Economics and Statistics, 91(4), 725-743.

Card, David, Alex Mas & Jesse Rothstein (2008) "Tipping and the Dynamics of Segregation," Quarterly Journal of Economics, 123 (1): 177-218.

Carrillo, Paul & Jonathan Rothbaum (2016) "Counterfactual Spatial Distributions," Journal of Regional Science, forthcoming.

Couture, Victor and Jessie Handbury (2016) "Urban Revival in America, 2000 to 2010," manuscript.

Damm, Anna Piil and Christian Dustmann (2014) "Does Growing Up in a High Crime Neighborhood Affect Youth Criminal Behavior?" American Economic Review, 104(6): 1806-32.

DiNardo, John, Nicole Fortin and Thomas Lemieux (1996) "Labor Market Institutions, and the Distribution of Wages 1973-1992: A Semiparametric Approach," Econometrica, 64, 1001-1044.

Edlund, Lena, Cecilia Machado & Michaela Sviatchi (2015) "Bright Minds, Big Rent: Gentrification and the Rising Returns to Skill," NBER Working Paper #21729.

Guerrieri, Veronica, Daniel Hartley and Erik Hurst (2013) “Endogenous Gentrification and Housing Price Dynamics,” Journal of Public Economics, 100, 45-60.

Hotz, Joseph & Miller (1994) "Conditional Choice Probabilities and the Estimation of Dynamic Models," The Review of Economic Studies 60(3), 497-529.

Jackson, Kenneth (1985) Crabgrass Frontier: The Suburbanization of the United States, Oxford University Press, New York.

Kain, John F. (1992) “The Spatial Mismatch Hypothesis: Three Decades Later,” Housing Policy Debate, 3:2, 371-462.

Lee, Sanghoon & Jeffrey Lin (2014) “Natural Amenities, Neighborhood Income Dynamics, and Persistence in the Spatial Income Distribution,” manuscript.

Margo, Robert (1992) "Explaining the postwar suburbanization of population in the United States: The role of income," Journal of Urban Economics 31(3): 301-310.

McKinnish, Terra, Randall Walsh & T. Kirk White (2010) "Who Gentrifies Low-Income Neighborhoods?" Journal of Urban Economics 67: 180-193.

Roback, Jennifer. 1982. “Wages, Rents, and the Quality of Life.” Journal of Political Economy, 90: 1257-1278.

Rosen, Sherwin. 1979. “Wages-based Indexes of Urban Quality of Life.” In Current Issues in Urban Economics, edited by P. Mieszkowski and M. Straszheim. Baltimore: Johns Hopkins Univ. Press.

Saiz, Albert (2010) “The Geographic Determinants of Housing Supply” Quarterly Journal of Economics, 125(3):1253-1296.

Table 1: Share of Population within 4 km of CBD in Tract Changing by at Least

	20 Percentile Points		1/2 Standard Deviation	
	up	down	up	down
	Panel A: Fraction White			
1970-1980	6.5%	13.3%	14.5%	20.8%
1980-1990	4.4%	6.0%	8.1%	13.9%
1990-2000	4.0%	3.1%	12.1%	11.0%
2000-2010	5.2%	1.3%	14.2%	5.5%
1980-2010	5.3%	1.3%	34.8%	23.2%
	Panel B: Fraction College Educated			
1970-1980	10.3%	10.0%	14.7%	7.6%
1980-1990	5.2%	5.8%	6.0%	7.5%
1990-2000	3.8%	6.1%	5.5%	7.6%
2000-2010	10.3%	4.0%	14.4%	5.3%
1980-2010	10.8%	4.0%	18.8%	16.6%
	Panel C: Median Income			
1970-1980	0.8%	12.3%	3.3%	21.3%
1980-1990	3.5%	1.1%	7.8%	3.3%
1990-2000	3.3%	1.4%	7.6%	2.9%
2000-2010	8.1%	1.5%	14.6%	4.7%
1980-2010	8.0%	1.4%	30.6%	9.1%

Notes: Distributions are within each of the 120 CBSAs in our sample. Each tract is weighted by its share of CBSA population.

Table 2: Decomposition of Percent Changes in Population within 2 and 4 km of CBDs
Based on Joint Population Distributions of Education and Race

					Contribution to Difference Between (1) and (2) from					
					Δchoices of				Δshares of	
Choices in year t	All	None	All	None	College+	College+	< College	< College	Educ|Race	Race
Shares in year t	All	None	None	All	White	NonWhite	White	NonWhite		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
CBD Radius										
Panel A: 1980-2000										
2 km	-0.07	0.21	-0.12	0.31	-0.01	0.00	-0.14	-0.18	-0.04	0.10
					(0.07)	(0.01)	(0.52)	(0.40)		
4 km	-0.07	0.21	-0.12	0.28	-0.02	-0.01	-0.16	-0.15	-0.04	0.09
					(0.07)	(0.01)	(0.56)	(0.36)		
Panel B: 2000-2010										
2 km	0.06	0.07	0.04	0.09	0.04	0.00	0.02	-0.08	-0.01	0.03
					(0.11)	(0.03)	(0.40)	(0.46)		
4 km	-0.01	0.07	-0.03	0.08	0.01	0.00	-0.02	-0.09	-0.01	0.03
					(0.11)	(0.03)	(0.42)	(0.44)		

Notes: All results are averages over the 120 CBSAs in our sample weighting each CBSA equally. Results in (1) and (2) report actual percent changes in population in the indicated CBD distance ring and average CBSA population growth rates respectively. Results in remaining columns use counterfactual data. See the text for a full explanation of the construction of each counterfactual. Table A1 presents the mathematical expression for each one. Results in (5)-(10) sum to actuals in (1) minus CBSA growth in (2). Entries in parentheses show the average fraction CBSA population in the indicated group.

Table 3: Decomposition of Percent Changes in Population within 2 km of CBDs for Additional Demographic Categories

					Contribution to Difference Between (1) and (2) from					
					∆choices of				∆shares of	
Choices in year t	All	None	All	None	Target	Target	NonTarget	NonTarget	X|Race	Race
Shares in year t	All	None	None	All	White	NonWhite	White	NonWhite		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Data Set										
Panel A: 1980-2000										
Age	-0.07	0.21	-0.14	0.34	0.01	-0.04	-0.15	-0.17	-0.03	0.10
					(0.19)	(0.11)	(0.40)	(0.30)		
Family Type	-0.07	0.21	-0.27	0.43	-0.11	-0.06	-0.12	-0.19	0.10	0.10
					(0.28)	(0.12)	(0.31)	(0.29)		
Income	-0.11	0.27	-0.19	0.37	0.00	-0.01	-0.24	-0.21	0.00	0.09
					(0.08)	(0.02)	(0.56)	(0.33)		
Panel B: 2000-2010										
Age	0.06	0.07	0.03	0.12	0.04	-0.01	0.01	-0.08	0.00	0.03
					(0.17)	(0.13)	(0.34)	(0.36)		
Family Type	0.05	0.08	-0.01	0.15	0.02	-0.03	-0.01	-0.08	0.03	0.03
					(0.27)	(0.15)	(0.24)	(0.34)		
Income	0.05	0.08	0.03	0.11	0.03	0.00	0.00	-0.08	0.00	0.02
					(0.09)	(0.03)	(0.47)	(0.40)		

Notes: Entries are analogous to those in Table 2 except that they are calculated using joint distributions of age, family type or income and race rather than education and race. The income joint distribution uses households rather than people. Target groups are ages 20-34, single or married without children and in the top 3 deciles of the sample area household income distribution.

Table 4: Decompositions of Changes in Demographic Composition within 2 km of CBDs

					Contribution to All in (1) from					
					Δchoices of				Δshares of	
Choices in year t	All	None	All	None	Target	Target	NonTarget	NonTarget	X|Race	Race
Shares in year t	All	None	None	All	White	NonWhite	White	NonWhite		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Outcome										
Data Set										
Panel A: 1980-2000										
Fraction White										
1 Education	-0.08	0.00	0.02	-0.11	-0.00	0.00	-0.05	0.08	0.01	-0.11
2 Income	-0.08	0.00	0.02	-0.10	0.00	0.00	-0.09	0.10	0.00	-0.10
Fraction College										
3 Education	0.06	0.00	0.01	0.05	-0.01	0.00	0.01	0.01	0.06	-0.01
Median Income (Percentile Points of Sample Area Distribution)										
4 Income	1.18	0.00	1.65	-0.23	0.08	-0.22	0.77	1.01	0.46	-0.93
Panel B: 2000-2010										
Fraction White										
1 Education	0.03	0.00	0.06	-0.04	0.02	-0.00	0.01	0.04	0.00	-0.04
2 Income	0.03	0.00	0.06	-0.02	0.01	0.00	0.00	0.04	0.00	-0.03
Fraction College										
3 Education	0.06	0.00	0.03	0.02	0.03	0.00	-0.01	0.01	0.03	-0.00
Median Income (Percentile Points of Sample Area Distribution)										
4 Income	3.84	0.00	4.19	-0.17	1.81	0.03	1.27	1.08	0.07	-0.42

Notes: Entries are analogous to those in Tables 2 and 3 except that the CBSA level statistic of interest differs. See the notes to Table 3 for a description of target groups and Table A1 for mathematical expressions used to calculate these counterfactuals.

Table 5: SES Index Regressions

	1970-1980	1980-1990	1990-2000	2000-2010	1980-2010
Estimator	RF	RF	IV	IV	RF
Panel A: Difference Specification					
1(< 4 km to CBD)	-0.219	0.008	0.009	0.143	0.174
	(0.021)	(0.013)	(0.008)	(0.008)	(0.042)
Standardized CBSA Emp Growth	-0.034	0.017	0.093	0.028	0.108
X 1(< 4 km to CBD)	(0.017)	(0.016)	(0.090)	(0.020)	(0.044)
Standardized CBD Area Emp Growth	0.059	0.009	0.054	0.116	0.055
X 1(< 4 km to CBD)	(0.014)	(0.015)	(0.043)	(0.035)	(0.037)
Observations	37,924	38,329	38,275	38,249	38,279
R-Squared (First Stage F)	0.112	0.038	(23.6)	(91.2)	0.140
Panel B: AR(1) Specification					
1(< 4 km to CBD)	-0.320	-0.045	-0.030	0.142	0.031
	(0.024)	(0.016)	(0.009)	(0.009)	(0.046)
Standardized CBSA Emp Growth	-0.033	0.014	0.136	0.026	0.112
X 1(< 4 km to CBD)	(0.018)	(0.019)	(0.093)	(0.024)	(0.047)
Standardized CBD Area Emp Growth	0.065	0.012	0.058	0.116	0.065
X 1(< 4 km to CBD)	(0.018)	(0.017)	(0.041)	(0.039)	(0.039)
Observations	37,924	38,329	38,306	38,281	38,279
R-Squared (First Stage F)	0.821	0.899	(22.7)	(103.5)	0.689
Panel C: Arrelano-Bond Specification					
1(< 4 km to CBD)		0.109	0.051	0.082	0.387
		(0.021)	(0.035)	(0.028)	(0.063)
Standardized CBSA Emp Growth		0.026	0.065	0.038	0.079
X 1(< 4 km to CBD)		(0.017)	(0.056)	(0.018)	(0.060)
Standardized CBD Area Emp Growth		0.007	0.024	0.042	0.062
X 1(< 4 km to CBD)		(0.018)	(0.033)	(0.025)	(0.064)
Observations		37,893	37,903	37,891	37,870

Notes: Each column in each panel reports results from a separate regression of the change in (Panel A) or level of (Panels B and C) the tract SES index on variables listed at left and indicators for 4-8, and 8-12 km from a CBD and 0-4, 4-8 and 8-12 km from the nearest top 1970 quartile SES index tract. Log of distance to the nearest coastline, lake, and river are also included as controls. Panel C implements an Arellano-Bond (1991) correction for endogeneity of the AR(1) variable in short panels. See Equations (10) , (11) and (12) in the text for specifications used in Panels A, B and C respectively. Employment growth variables and their Bartik instruments are standardized to be mean 0 and standard deviation 1. "RF" refers to "reduced form" and "IV" stands for "instrumental variables" in column headers. Observations are weighted such that each CBSA and 0-4 km, 4-8 km, 8-12 km and beyond CBD distance rings within each CBSA receive equal weight . Coefficients that are significant at the 10% level are shaded red if positive and blue if negative. RF standard errors are clustered by CBSA.

Table 6: Changes in Tract Valuations by Race and Education

Estimator	1980-1990 RF	1990-2000 IV	2000-2010 IV
Panel A: White College+			
1(<4 km to CBD)	-0.064	-0.229	0.100
	(0.067)	(0.295)	(0.058)
CBSA Employment Growth* 1(< 4 km to CBD)	-0.170	0.066	0.007
	(0.045)	(0.209)	(0.055)
CBD Area Employment Growth * 1(< 4 km to CBD)	0.039	0.094	0.286
	(0.027)	(0.103)	(0.141)
Observations	32,712	32,712	32,712
R-Squared (First Stage F)	0.074	(18.5)	(38.9)
Panel B: Black College+			
1(<4 km to CBD)	-0.481	-1.129	0.159
	(0.283)	(0.742)	(0.134)
CBSA Employment Growth* 1(< 4 km to CBD)	-0.068	0.472	-0.252
	(0.171)	(0.530)	(0.104)
CBD Area Employment Growth * 1(< 4 km to CBD)	-0.007	-0.202	0.261
	(0.059)	(0.252)	(0.182)
Observations	14,413	14,413	14,413
R-Squared (First Stage F)	0.085	(7.5)	(48.9)
Panel C: White <HS			
1(<4 km to CBD)	-0.223	-0.070	0.007
	(0.048)	(0.287)	(0.060)
CBSA Employment Growth* 1(< 4 km to CBD)	-0.086	-0.062	-0.122
	(0.033)	(0.203)	(0.055)
CBD Area Employment Growth * 1(< 4 km to CBD)	0.015	0.075	0.239
	(0.016)	(0.095)	(0.148)
Observations	33,301	33,301	33,301
R-Squared (First Stage F)	0.108	(19.1)	(50.4)
Panel D: Black <HS			
1(<4 km to CBD)	-0.020	0.409	0.095
	(0.176)	(0.594)	(0.135)
CBSA Employment Growth* 1(< 4 km to CBD)	-0.258	-0.597	-0.338
	(0.106)	(0.423)	(0.124)
CBD Area Employment Growth * 1(< 4 km to CBD)	-0.019	0.207	0.298
	(0.037)	(0.216)	(0.215)
Observations	13,625	13,625	13,625
R-Squared (First Stage F)	0.115	(10.0)	(51.0)

Notes: Reported coefficients are from regressions analogous to those in Table 5 Panel A, except using changes in λ utility components for each group indicated in panel headers rather than the unified SES index, with group specific samples defined so as to be consistent over the full 1980-2010 period. Equation (13) in the text shows the full regression specification used. CBSA and CBD area employment shocks are normalized to be mean 0 and standard deviation 1. All regressions have CBSA fixed effects. Weights are the same as in Table 4. Coefficients that are significant at the 10% level are shaded red if positive and blue if negative.

Table 7: Contributions to Changes in Central Area Population Growth by Various Demographic Groups Using the Model

	Within 2 km of CBDs				Within 4 km of CBDs			
	College White	College NonWhite	<College White	<College NonWhite	College White	College NonWhite	<College White	<College NonWhite
Due to ...								
Panel A: 1980-2000								
Chg. In Home Prices	0.000	0.000	0.002	-0.006	0.000	0.000	0.002	-0.002
Chg. In Central Emp. Opp.	-0.001	0.001	-0.024	0.016	-0.002	0.001	-0.027	0.014
Income Effect	-0.005	0.002	-0.099	-0.062	-0.005	0.003	-0.110	-0.051
Val. Of Observed Amenities	0.001	0.000	-0.017	-0.003	0.000	0.000	-0.018	-0.001
Unobserved Amenities	-0.012	-0.009	-0.026	-0.128	-0.013	-0.009	-0.027	-0.118
Unexplained	0.005	0.000	0.022	-0.017	0.003	0.000	0.022	0.010
Total	-0.01	0.00	-0.14	-0.20	-0.02	0.00	-0.16	-0.15
Panel B: 2000-2010								
Chg. In Home Prices	0.001	0.001	0.001	-0.005	0.000	0.000	0.000	-0.004
Chg. In Central Emp. Opp.	0.004	0.000	0.003	0.009	0.004	0.000	0.001	0.010
Income Effect	0.001	-0.005	-0.009	-0.044	0.001	-0.005	-0.009	-0.044
Val. Of Observed Amenities	0.000	-0.001	-0.010	-0.003	-0.002	-0.001	-0.010	-0.003
Unobserved Amenities	0.013	0.005	0.021	-0.021	0.012	0.005	0.022	-0.023
Unexplained	0.009	-0.001	-0.002	-0.020	-0.012	-0.002	-0.026	-0.018
Total	0.03	0.00	0.00	-0.08	0.00	0.00	-0.02	-0.08

Notes: Each entry is the marginal contribution of the component listed at left on central area population within the CBD distance ring indicated at top because of shifts in neighborhood choices of the demographic group indicated at top. Columns do not always sum exactly to entries in Table 2 because of minor sample differences, as is explained in the text.

Table A1: Explanation of Counterfactual Experiments
Population Distributions Used to Construct Counterfactuals

Column in Tables 2-3	Choices	Shares	Group: Race	Group: X-Dimension	Math Notation
1	All t	All t	All	All	$f_{jt}(i\|r,x)g_{jt}(r,x)$
2	All Base Yr	All Base Yr	All	All	$f_{jb}(i\|r,x)g_{jb}(r,x)$
3	All t	All Base Yr	All	All	$f_{jt}(i\|r,x)g_{j8}(r,x)$
4	All Base Yr	All t	All	All	$f_{j8}(i\|r,x)g_{jt}(r,x)$
5	Target Whites t	All Base Yr	Whites	Target	$f_{jt}(i\|r,x)g_{j8}(r,x)$
			Blacks, Others	Target	$f_{j8}(i\|r,x)g_{j8}(r,x)$
			Whites	Non-Target	$f_{j8}(i\|r,x)g_{j8}(r,x)$
			Blacks, Others	Non-Target	$f_{j8}(i\|r,x)g_{j8}(r,x)$
6	Target t	All Base Yr	Whites	Target	$f_{jt}(i\|r,x)g_{j8}(r,x)$
			Blacks, Others	Target	$f_{jt}(i\|r,x)g_{j8}(r,x)$
			Whites	Non-Target	$f_{j8}(i\|r,x)g_{j8}(r,x)$
			Blacks, Others	Non-Target	$f_{j8}(i\|r,x)g_{j8}(r,x)$
7	Target+Whites t	All Base Yr	Whites	Target	$f_{jt}(i\|r,x)g_{j8}(r,x)$
			Blacks, Others	Target	$f_{jt}(i\|r,x)g_{j8}(r,x)$
			Whites	Non-Target	$f_{jt}(i\|r,x)g_{j8}(r,x)$
			Blacks, Others	Non-Target	$f_{j8}(i\|r,x)g_{j8}(r,x)$
8	All t	All Base Yr	All	All	$f_{jt}(i\|r,x)g_{j8}(r,x)$
9	All t	X|r in t, r in Base Yr	All	All	$f_{jt}(i\|r,x)g_{jt}(x\|r)h_{j8}(r)$
10	All t	All t	All	All	$f_{jt}(i\|r,x)g_{jt}(x\|r)h_{jt}(r)$

Notes: Entries in the final column show the contribution of each demographic group to each counterfactual in Tables 2-3. See Section 3.1 of the text for an explanation of notation. Target groups are college graduates, households in the top three deciles of the income distribution, people aged 20-34 and singles or married couples with no kids.

Table A2: Aggregate Quantities

	Fraction White	Fraction College	Median HH Income	Share in Families without Kids	Share 20-34
Panel A: Entire Sample					
1970	0.883	0.116	47881		
1980	0.836	0.102	44266	0.328	0.266
1990	0.809	0.138	52310	0.357	0.255
2000	0.753	0.167	58308	0.384	0.211
2010	0.717	0.196	55532	0.401	0.209
Panel B: Within 2 km of CBDs					
1970	0.683	0.082	32626		
1980	0.590	0.085	26281	0.404	0.300
1990	0.548	0.115	30991	0.376	0.317
2000	0.507	0.144	36770	0.420	0.298
2010	0.533	0.204	38423	0.454	0.324
Panel C: Within 4 km of CBDs					
1970	0.722	0.089	36523		
1980	0.629	0.087	31055	0.366	0.288
1990	0.584	0.115	35777	0.358	0.289
2000	0.531	0.139	40934	0.396	0.267
2010	0.537	0.183	39882	0.423	0.286

Notes: Each entry is an average across CBSAs in the sample.

Table A3: Decomposition of Percent Changes in Population - Reverse Order

	Contribution to Difference Between (1) and (2) in Tables 2 and 3					
	from Δshares of		from Δchoices of			
Choices in year t	X|Race	Race	Target	Target	NonTarget	NonTarget
Shares in year t			White	NonWhite	White	NonWhite
	(1)	(2)	(3)	(4)	(5)	(6)
Data Set & CBD Distance Ring						
Panel A: 1980-2000						
Education, 2km	-0.04	0.13	-0.02	-0.01	-0.11	-0.24
Education, 4km	-0.03	0.10	-0.03	-0.01	-0.12	-0.19
Age, 2km	0.00	0.13	0.01	-0.04	-0.14	-0.23
Family Type, 2km	0.10	0.12	-0.11	-0.09	-0.09	-0.21
Income, 2km	0.00	0.10	0.00	-0.01	-0.20	-0.27
Panel B: 2000-2010						
Education, 2km	-0.02	0.05	0.04	0.00	0.02	-0.09
Education, 4km	-0.02	0.04	0.01	0.00	-0.01	-0.09
Age, 2km	0.01	0.05	0.04	-0.01	0.01	-0.09
Family Type, 2km	0.03	0.04	0.02	-0.03	-0.01	-0.09
Income, 2km	0.00	0.03	0.03	0.00	0.00	-0.09

Notes: Results are analogous to those in Tables 2 and 3. The only difference is the order in which the counterfactuals are imposed.

Table A4: Descriptive Statistics for Employment Shocks

Panel A: Employment Shocks

	Δ ln(CBSA Employment)			Δ ln(Employment Within 4 km of CBD)		
	Mean	SD	Coeff of Var	Mean	SD	Coeff of Var
1980-1990	0.17	0.12	1.42		Not Available	
1990-2000	0.10	0.09	1.11	-0.07	0.12	-0.58
2000-2010	0.08	0.09	0.89	-0.01	0.13	-0.08

Panel B: Instruments

	Bartik			Spatial Bartik		
	Mean	SD	Coeff of Var	Mean	SD	Coeff of Var
1970-1980	0.11	0.02	5.15	0.14	0.02	6.29
1980-1990	0.17	0.03	5.99	0.20	0.02	8.27
1990-2000	0.05	0.03	1.49	0.10	0.03	3.00
2000-2010	0.07	0.03	2.44	0.08	0.02	3.54
1980-2010	0.29	0.08	3.64	0.39	0.07	5.23

Notes: We only use actual employment shocks for the 1990-2000 and 2000-2010 periods in Tables 5, 6 and 7, instrumented with variables whose summary statistics are reported in Panel B. For other periods, those tables report reduced form results. Statistics are for the 120 CBSAs in the sample.

Table A5: Patterns of Housing Costs in Tracts within 4 km of CBDs

	1970-1980	1980-1990	1990-2000	2000-2010	1980-2010
Estimator	RF	RF	IV	IV	RF
Panel A: Difference Specification					
1(< 4 km to CBD)	-0.063	-0.028	-0.028	0.011	-0.001
	(0.016)	(0.012)	(0.007)	(0.008)	(0.021)
Standardized CBSA Emp Growth	-0.036	0.009	0.055	0.017	0.028
X 1(< 4 km to CBD)	(0.010)	(0.013)	(0.085)	(0.019)	(0.027)
Standardized CBD Area Emp Growth	0.032	0.045	0.066	0.084	0.082
X 1(< 4 km to CBD)	(0.014)	(0.016)	(0.044)	(0.041)	(0.028)
Observations	31,011	35,704	37,096	36,715	35,078
R-Squared (First Stage F)	0.490	0.626	(23.7)	(71.4)	0.423
Panel B: AR(1) Specification					
1(< 4 km to CBD)	-0.063	-0.007	0.006	0.036	0.040
	(0.015)	(0.012)	(0.007)	(0.008)	(0.022)
Standardized CBSA Emp Growth	-0.035	-0.008	-0.125	0.022	0.001
X 1(< 4 km to CBD)	(0.009)	(0.013)	(0.073)	(0.020)	(0.025)
Standardized CBD Area Emp Growth	0.037	0.027	0.135	0.031	0.074
X 1(< 4 km to CBD)	(0.013)	(0.013)	(0.039)	(0.040)	(0.026)
Observations	31,011	35,704	35,572	36,330	35,078
R-Squared (First Stage F)	0.046	0.026	(30.5)	(92.1)	0.039
Panel C: AR(1) Specification, Arellano-Bond Adjustment					
1(< 4 km to CBD)		-0.051	0.006	-0.064	0.048
		(0.040)	(0.030)	(0.071)	(0.052)
Standardized CBSA Emp Growth		-0.005	0.005	-0.024	0.026
X 1(< 4 km to CBD)		(0.028)	(0.025)	(0.031)	(0.052)
Standardized CBD Area Emp Growth		0.049	0.036	0.044	0.106
X 1(< 4 km to CBD)		(0.029)	(0.051)	(0.046)	(0.071)
Observations		30,944	35,450	36,144	30,432

Notes: Each column in each panel reports results from a separate regression of the change in tract owner occupied housing price index using the same specification as in Table 6. The housing cost index is formed from the residuals of a regression of log mean owner occupied home value on housing unit structure characteristics (number of units in building, number of bedrooms in unit, age of building) of the tract and CBSA fixed effects. Coefficients that are significant at the 10% level are shaded red if positive and blue if negative. Observations are weighted by their CBSA share, with areas within 4 km of CBSAs weighted equally.

Figure 1: Share Within 4 km of the CBD in a Top Half SES Distribution Census Tract

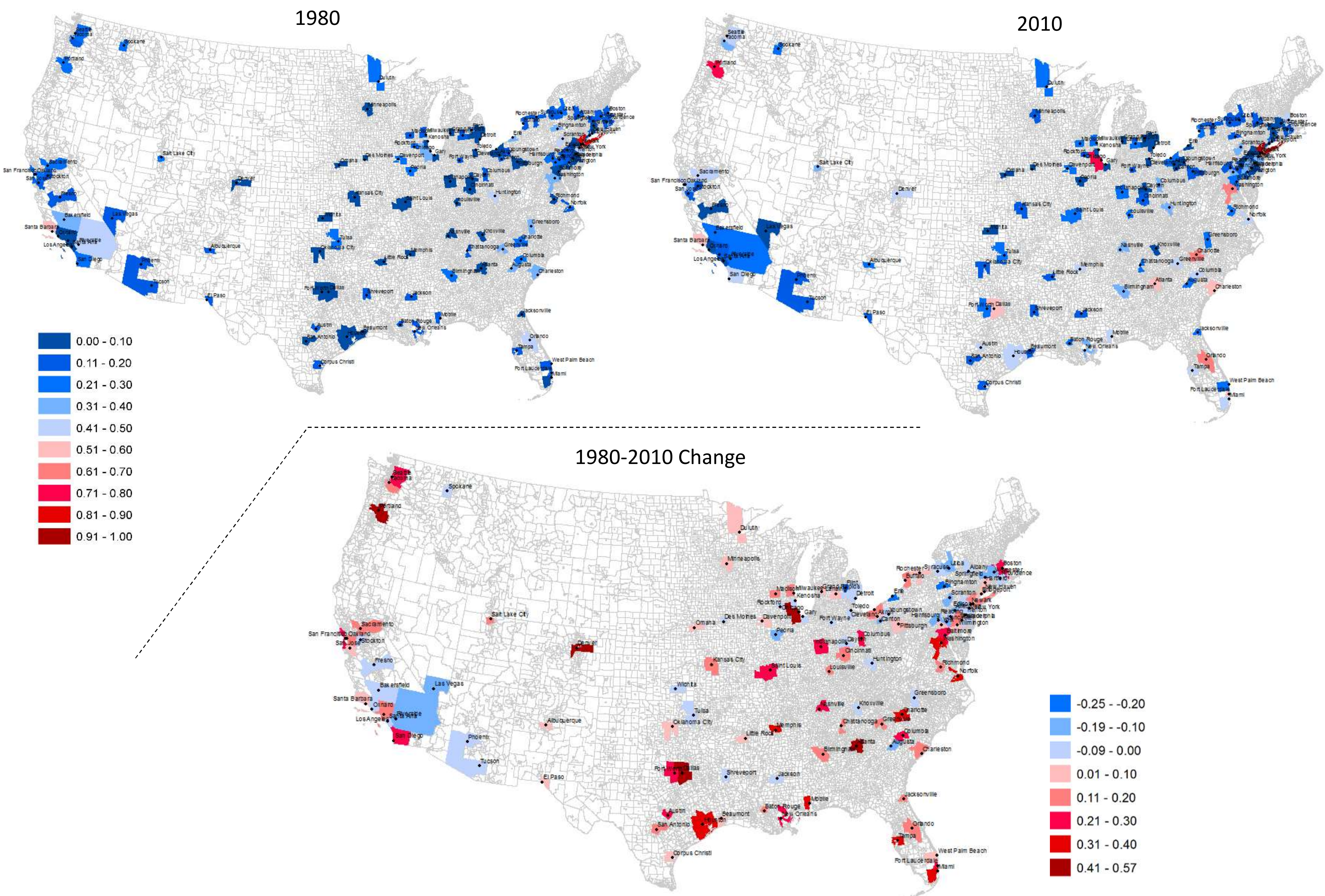
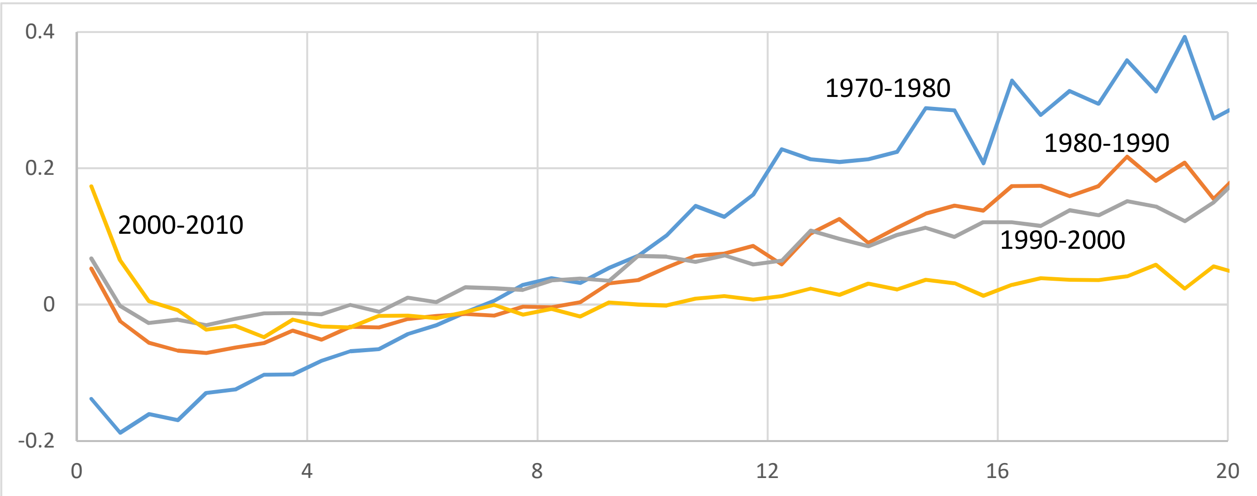


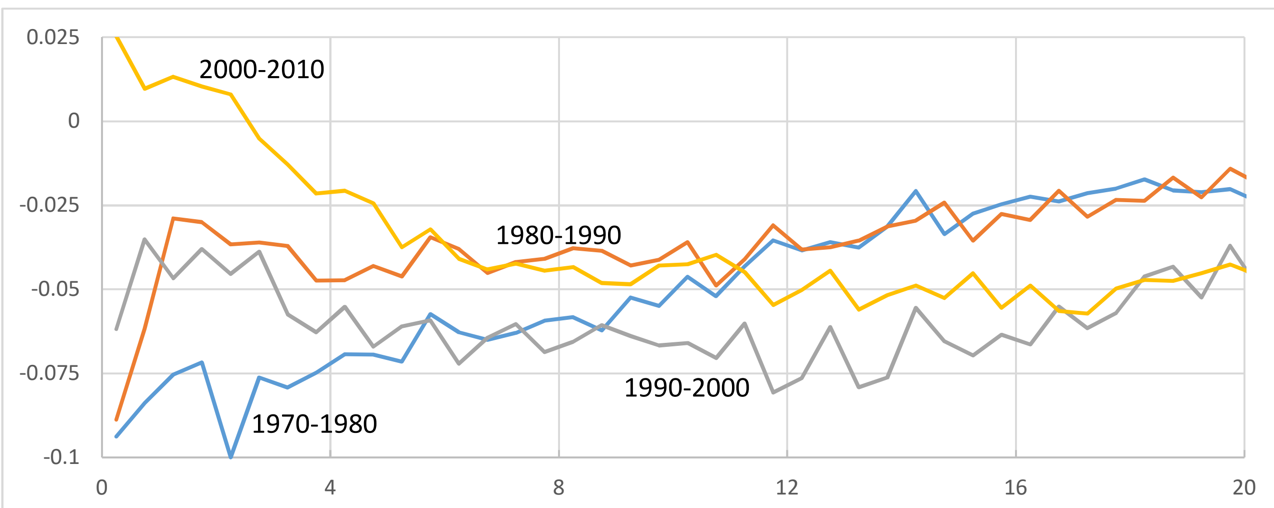
Figure 2: Measures of Gentrification as a Function of CBD Distance (km)

Medians Across 120 CBSAs, 0.5 km CBD Distance Bands

Panel A: Percent Change in Population



Panel B: Change in Fraction White



Panel C: Change in Fraction 25+ with College Education

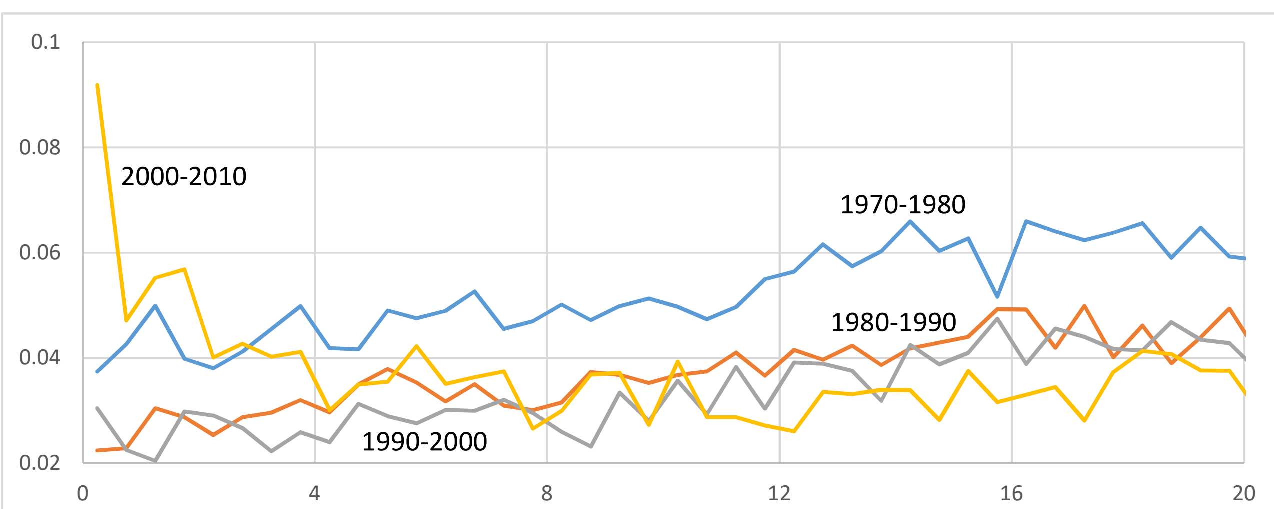


Figure 3: 1980-2010 Neighborhood Change in Chicago

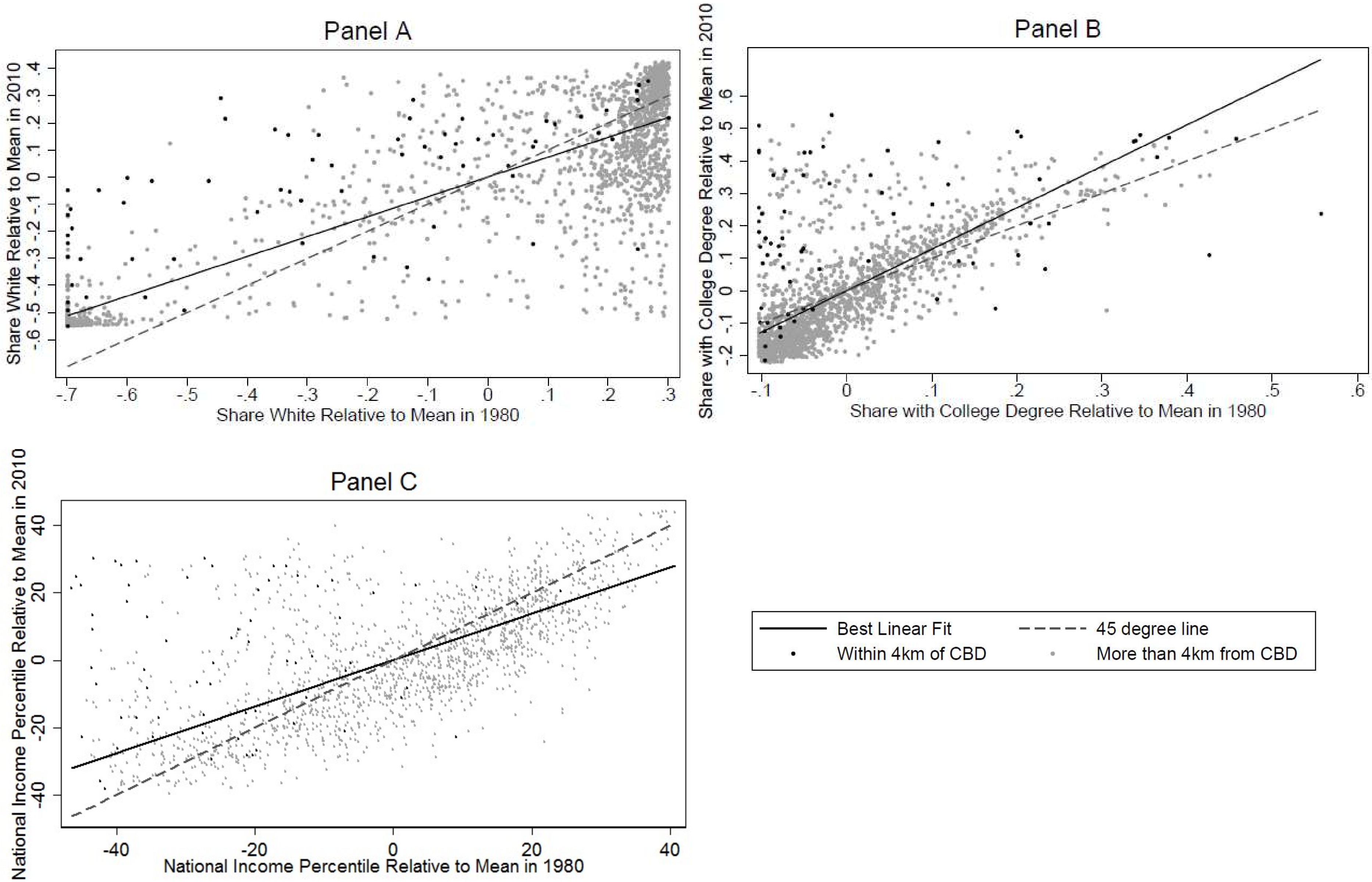
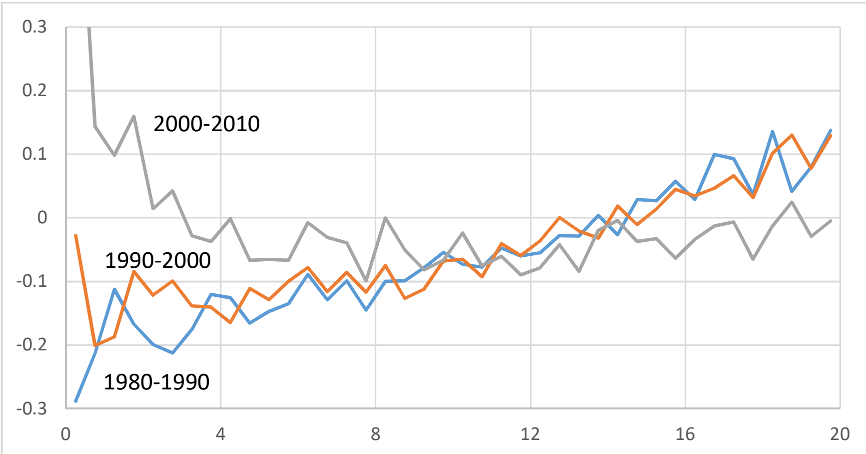
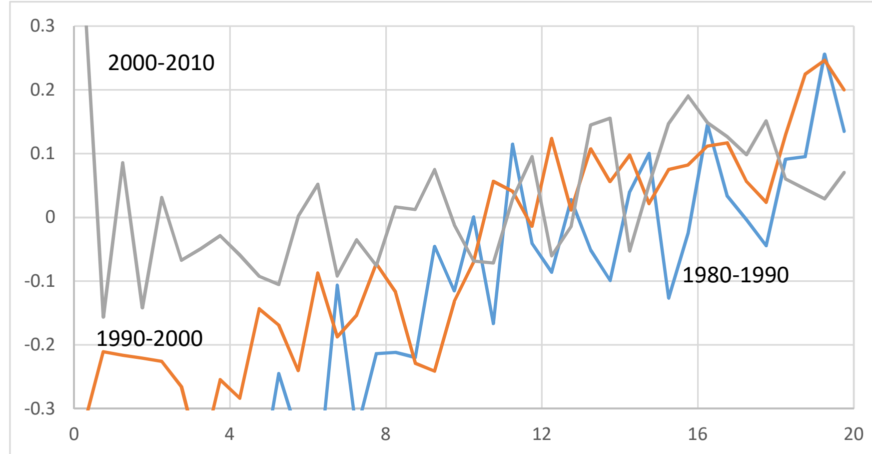


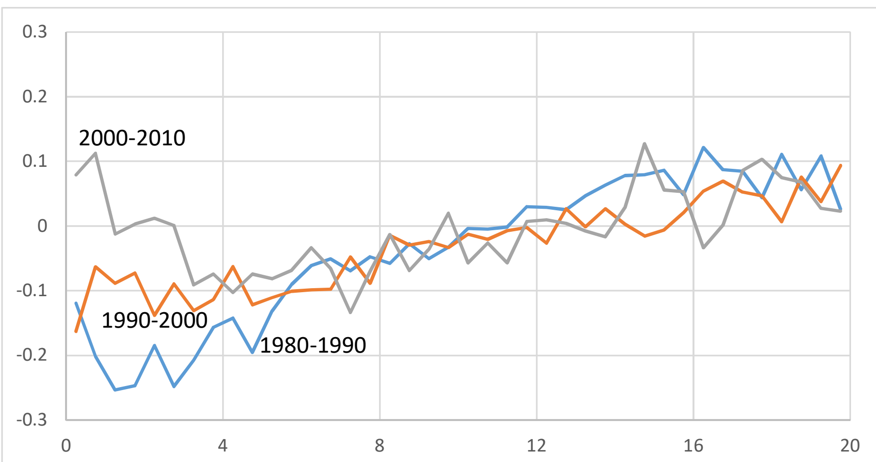
Figure 4: Changes in Neighborhood Valuations as a function of CBD Distance by Race and Education



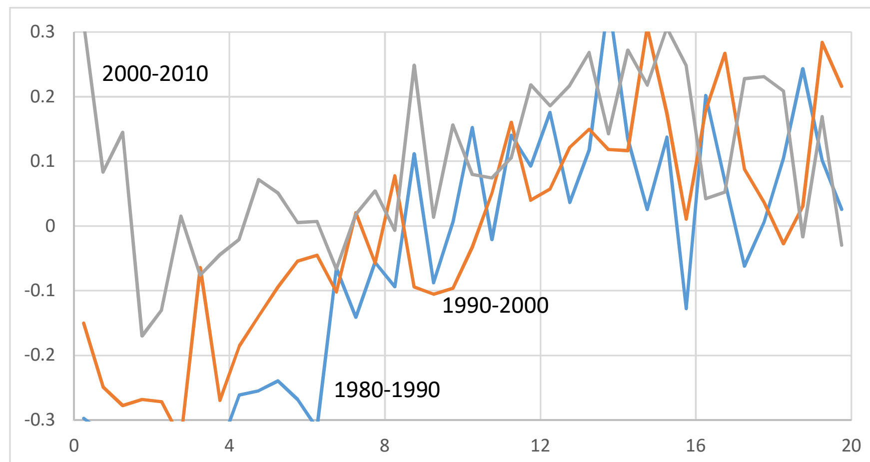
Panel A: Whites with College or More



Panel B: Blacks with College or More



Panel C: White High School Dropouts



Panel D: Black High School Dropouts

Notes: Each plot indicates the average change across CBSAs in λ for the indicate demographic group over the indicated decade. λ is calculated for each 0.5 km CBD distance band using the expression in Equation (5) in the text.