Quantifying the Welfare Effects of Neighborhood Change on Incumbent Low-Income Renters

Robert French, Ashvin Gandhi, Valentine Gilbert

November 6, 2023

Please read the latest draft here.

How does gentrification affect the incumbent residents of low-income neighborhoods? This paper investigates whether changes in neighborhood amenities and employment access accompanying gentrification are sufficient to compensate incumbent renters for increased housing costs. We link person-level administrative US Census data from the American Community Survey, the Master Address File, and the Longitudinal Employer-Household Dynamics database to construct an annual panel that tracks the earnings, workplaces, and residential addresses of over 1 million low-income urban renter households through 2000-2019. We use these data to estimate a dynamic structural model of residential and workplace choice. We identify our model with skill-specific labor demand shocks to potential commuting destinations constructed using geocoded data on the near universe of US business establishments from the Census Bureau's Longitudinal Business Database. Our estimates suggest that incumbent low-income renters benefited on average from gentrification relative to a hypothetical steady state equilibrium formed using the observed spatial distribution of neighborhood rents, college graduates, and jobs in the year 2000. Most of the variation in welfare effects is between metro areas, indicating that where low-income renters lived within US metros mattered far less than which US metro they lived in.

Acknowledgments

1. Introduction

Gentrification is associated with simultaneous increases in housing costs and changes in both public (e.g. schools and public safety) and private (e.g. restaurants and bars) neighborhood amenities. Residents may benefit from gentrification if they value the change in amenities by more than the amount housing costs increase (Vigdor 2010). Conversely, residents may be harmed if changes in amenities are insufficient to compensate for rising housing costs. Residents with strong attachments to their home neighborhood are especially vulnerable to rising housing costs as they are less willing to relocate given any cost increase. This paper leverages extensive US Census Bureau data to quantify these tradeoffs for incumbent renters in gentrifying neighborhoods throughout 2000-2019.[1]

Beginning in the 1990s and intensifying after the year 2000, gentrification transformed the socioeconomic composition of vast areas within American inner cities (Couture and Handbury 2023).[2] Between 2000 and 2017, Census tracts closest to their metros' central business districts (CBDs) increased their share of residents with a college degree by an average of 15-percentage points from a base of 24 percent. This contrasts with a 7-percentage point average increase in suburban neighborhoods' shares of college graduates, reflecting a secular increase in educational attainment.[3] How this transformation of American inner cities affected the incumbent residents of gentrifying areas remains an open question.

The gentrification of inner-city America reversed postwar urban decline, wherein middle- and upper-class households left the inner city in favor of suburban life (Jackson (1987), Boustan (2010), Mieszkowski and Mills (1993)). This postwar suburbanization is considered a major contributor to concentrated inner-city disadvantage (Wilson (1987)). At the same time primarily white middle- and upper-class households were leaving to the suburbs, inner-city violent crime increased ten-fold (Cullen and Levitt (1999), Curci and Masera (2023)), city public finances declined (Derenoncourt 2022), and many employers relocated to the suburbs (Kain (1968),

[1]Our focus on low-income renter households stems from their vulnerability to the financial costs of gentrification and the fact that in 2010, 64% of urban housing units occupied by households with incomes below $50,000 were rented (Manson et al. 2022). Recent research does show the potential for homeowners to be harmed from gentrification due to rising property taxes (Ding and Hwang (2020), Berry (2021), Fu (2022)). Evaluating the welfare effects of gentrification on incumbent homeowners is an interesting question for future research.

[2]The economic forces causing demand for inner-city living to rise among college graduates were multifaceted. Rising top incomes increased the demand for local service amenities concentrated in downtown neighborhoods (Couture et al. 2023) and raised the time-cost of commuting (Edlund, Machado, and Sviatschi (2022), Su (2022)). Declining urban crime (Ellen and O'Regan (2010), Ellen, Horn, and Reed (2019)), shifting preferences for urban amenities (Brueckner, Thisse, and Zenou (1999), Glaeser, Kolko, and Saiz (2001), Baum-Snow and Hartley (2020), Couture and Handbury (2020)), evolving transportation infrastructure (LeRoy and Sonstelie (1983), Glaeser, Kahn, and Rappaport (2008)), and delayed child bearing (Moreno-Maldonado and Santamaria 2022) all likely contributed to the demand for downtown living. These forces were compounded by increases in the valuation of downtown amenities *caused* by the growing presence of college graduates (Berkes and Gaetani (2023), Diamond (2016), Guerrieri, Hartley, and Hurst (2013)).

[3]See appendix A for more details on neighborhood change since the turn of the century.

Glaeser and Kahn (2001), Miller (2021)). If suburbanization was a significant factor behind these changes, its reversal over the past few decades might have caused environmental shifts favored by incumbent inner-city residents. Residents may have benefited from more local job opportunities, improved public amenities funded by rising land values, and greater access to private consumption amenities like grocery stores and restaurants.

Like any other good, though, improvements in neighborhood quality are invariably accompanied by rising prices. Incumbent residents could have been harmed by gentrification if housing costs increased by more than residents' willingness to pay for the accompanying amenity changes. That concerns over housing affordability are today ubiquitous in many large US cities makes this possibility salient (Glaeser (2020), Gabriel and Painter (2020)). High moving costs, neighborhood-specific accumulated capital, and sparse choice sets may also have exacerbated the potential harms of gentrification. By rendering incumbent residents less responsive to changing neighborhood characteristics, these moving frictions ensure incumbent residents bear the incidence of gentrification.[4]

To quantify the welfare effects of gentrification on incumbent residents, this paper estimates a dynamic model of residential and workplace choice that accounts for welfare-relevant neighborhood characteristics, a rich set of moving frictions, and changes in city-wide neighborhood choice sets. With our parameter estimates in hand, we explore how incumbent residents were affected by gentrification between 2000 and 2019. Specifically, we ask whether incumbent renters of low-income neighborhoods benefited from gentrification relative to a hypothetical steady state equilibrium formed using the observed spatial distribution of neighborhood rents, college graduates, and jobs in the year 2000. We finally conduct experiments to measure how various neighborhood and CBSA-level characteristics influence the impact of gentrification on incumbent renters.

We conduct our analysis on de-identified person-level data from the US Census Bureau's Master Address File (MAF), which records the near universe of US adults' residential migration histories from 2000 onward. We link these data to persons' earnings and workplace locations from the employer-employee linked Longitudinal Employer-Household Dynamics (LEHD) database, which records the near universe of private sector, state, and local government workers' employment histories between 2000-2019. We finally link these data to person-level sociodemographic information from all American Community Survey (ACS) respondents (2005-2021), as well as property-level data from the Census Bureau's Master Address File Extract (MAF-X) and CoreLogic's residential property databases (2006-2019). Together, these data provide us with the distinct ability to observe the residential locations, earnings, and workplaces of over 1 million

[4]Incumbent residents may also be harmed if the supply of local amenities is skewed away from their preferences and toward those of the gentrifiers (Almagro and Domínguez-Lino 2022), or if gentrification alters neighborhoods' racial composition away from the preferences of incumbents (Davis, Gregory, and Hartley 2023).

low-income urban renter households across 50 large metros for up to 20 years during the most intense period of recent gentrification.

Our data yield three facts. First, low-income renter households are highly mobile. Only 54% of our sample resided in the same Census tract in 2013 as they did in 2010. However, the likelihood a renter household leaves its origin Census tract decreases significantly with longer tenure. Of renter households staying in their origin tract between 2010 and 2013, 77% remain there for the next three years. Second, gentrification is not associated with changes in the earnings, commutes, or residential tenures for our sample of incumbent renter households. This is true across a range of baseline neighborhood environments. Third, despite low-income renters being highly mobile, gentrification does affect the neighborhood characteristics incumbent renters experience. A one standard deviation increase in our measure of gentrification is associated with a 7.7% (5.4%) increase in the neighborhood rents and a 33.2% (36.6%) increase in the neighborhood college share incumbent Black (Non-Black) renters in poor neighborhoods experience between 2010 and 2019.[5] These effect sizes suggest the potential for large welfare effects from gentrification among low-income incumbent renters.

In our dynamic residential and workplace choice model, heterogeneous agents choose their neighborhood and workplace locations each period to maximize their expected lifetime utility. Agents are subject to a rich set of moving costs, can accumulate neighborhood-specific capital, and are forward-looking. They obtain flow utilities comprised of expected housing and non-housing consumption, neighborhood amenities, and their accumulated neighborhood capital. Conditional on neighborhood rents, agents' expected consumption varies across neighborhoods because of differences in their commute-time-discounted proximity to jobs. This feature of our model captures the fact that neighborhoods farther from employment centers are less desirable due to the increased financial cost of commuting (Le Barbanchon, Rathelot, and Roulet 2021).

We model neighborhood amenities as a function of neighborhoods' shares of college graduates. This choice is motivated by a literature documenting a robust positive relationship between the provision of local public and private amenities and the local share of college graduates (Glaeser, Kolko, and Saiz (2001), Diamond (2016), Autor, Palmer, and Pathak (2017), Su (2022), Almagro and Domínguez-Lino (2022), Hoelzlein (2023)). We further allow neighborhood amenities to vary over time due to unobserved factors not caused by changes in the local share of college graduates. While the evolution of these unobserved exogenous neighborhood amenities is not caused by shifts in the local college share, they may nonetheless be correlated with households' residential location choices, presenting a challenge to identification.

We identify our model parameters by combining establishment-level employment data

[5]Changes in "experienced" neighborhood characteristics refer to the differences in the characteristics of the neighborhoods residents reside in during 2019 relative to the characteristics of the neighborhoods residents reside in during 2010.

from the Census Bureau's Longitudinal Business Database (LBD) with advances in the quasi-experimental "shift-share" literature. Specifically, we construct two sets of instrumental variables to disentangle preferences for observed neighborhood characteristics and unobserved exogenous neighborhood amenities. The first set of instrumental variables aggregates skill-specific shocks to potential commuting destinations. The intuition behind these instruments is that as neighborhoods' access to high-skill employment opportunities improve, they become more desirable for college graduates and thus are more likely to gentrify all else equal. These sets of instrumental variables build on recent work in Baum-Snow, Hartley, and Lee (2019) and Baum-Snow and Han (2023) that microfound measures of employment access with commuting data in a workplace choice model á la Tsivanidis (2022). Our establishment-level business data make identification from national industry shocks plausible, allowing the location of business establishments to be correlated with unobserved neighborhood characteristics (Borusyak, Hull, and Jaravel 2022).

Our second set of instrumental variables are motivated by the observation that gentrification tends to occur near neighborhoods with already high shares of college graduates (Guerrieri, Hartley, and Hurst 2013). For each census tract in our data, we construct distance-weighted measures of proximity to other neighborhoods' shares of college graduates. We then exploit the fact that our analysis spans 50 metro areas, and interact our neighborhood-level proximity measures with CBSA-wide Bartik labor demand shocks. Identification then proceeds analogously to a difference-in-difference estimator: we compare differences in gentrification between neighborhoods near and far already gentrified tracts in CBSAs experiencing large labor demand shocks to differences in gentrification between neighborhoods near and far already gentrified tracts in CBSAs not experiencing large labor demand shocks.

Both Black and Non-Black low-income renter households in our sample have a positive valuation for the share of college graduates in a neighborhood, consistent with the existing literature. Black households, however, appear to place little weight on market access. This contrasts with Non-Black households who value residing near plentiful employment opportunities. Black low-income renter households find it especially costly to leave their home neighborhood; our estimates suggest the fixed cost of moving within one's own CBSA is over \$3,500 (\$1,600) for Black (Non-Black) households, and that this cost increases by around \$600 (\$250) with each additional year of residential tenure.

We finally ask what our parameter estimates imply about the welfare effects of gentrification for low-income incumbent renters during 2000-2019. Specifically, we run the following experiment. We calculate the expected welfare for hypothetical low-income renters initially residing in each urban low-income Census tract during the year 2000 under the assumption that the economy is in steady state. To do so, we find the continuation values and levels of exogenous neighborhood amenities that exactly rationalize a stationary equilibrium given the observed distribution neighborhood rents, college graduates, and jobs in the year 2000. We then recompute expected welfare for each hypothetical renter household using observed changes in these equilibrium objects between 2000 and 2019. Comparisons between these expected utilities yield neighborhood-level estimates of changes to incumbent renters' welfare post-2000. We find that, on average, incumbent renters originally living in poor neighborhoods experienced an increase in expected welfare due to gentrification. We are currently unpacking these results.

Relation to Literature. Our paper contributes to three strands of literature. First, we contribute to the literature on the welfare implications of spatial sorting. Grounded in the canonical spatial equilibrium models stemming from Rosen (1974) and Roback (1982), an empirical literature has sought to quantify the implications of urban spatial sorting for households differentiated by their educational attainment. Moretti (2013) studied the implication of cross-metro sorting for real income inequality, while Diamond (2016) incorporated the endogenous supply of citywide amenities to study the implications of cross-metro sorting for welfare inequality. Su (2022) and Couture et al. (2023) similarly examine the welfare implications of spatial sorting, but focus on within-metro sorting.[6] Other closely related papers focus on understanding the emergence of endogenously provided local amenities (Couture and Handbury (2020), Almagro and Domínguez-Lino (2022), Hoelzlein (2023), Glaeser, Luca, and Moszkowski (2023)). With few exceptions, the existing literature quantifies the effect of spatial sorting on the expected welfare of *prospective* city residents.[7] Our paper instead exploits rich panel data to examine the impact of gentrification on the welfare of *incumbent* renters. We show how strong neighborhood attachments, high moving costs, and location-dependent neighborhood choice sets mediate the welfare impacts of gentrification across heterogeneous neighborhood environments.[8]

Second, we contribute to the empirical residential choice literature that estimates households' willingness to pay for housing and neighborhood characteristics. Earlier static models of residential choice (Brock and Durlauf (2002), Bayer, Ferreira, and McMillan (2007), Vigdor (2010)) have given way to dynamic models that account for moving frictions and forward looking behavior (Kennan and Walker (2011), Bishop (2012), Bayer et al. (2016)). Researchers have used

[6]See Kuminoff, Smith, and Timmins (2013) and Diamond and Gaubert (2022) for a comprehensive review of these and other papers examining the implications of spatial sorting on inequality. This research in turn contributes more broadly to the quantitative spatial equilibrium literature summarized in Redding and Rossi-Hansberg (2017).

[7]Balboni et al. (2020) is a notable exception, which uses a repeated static commuting model coupled with the exact-hat algebra of Dekle, Eaton, and Kortum (2008) to estimate the welfare impacts of transit infrastructure investments and the resulting sorting of households on the welfare of incumbent residents in Dar es Salaam, Tanzania.

[8]Urban housing policies like rent control and eviction protections prioritize incumbent renters' welfare over landlords and residents unprotected by such policies (Glaeser and Luttmer (2003), Diamond, McQuade, and Qian (2018), Collinson et al. (2023), Abramson (2023)). Understanding the difference in welfare effects between prospective and incumbent residents is thus critical to discerning the appropriate set of policy responses to gentrification.

these dynamic neighborhood choice models to estimate preferences over the racial composition of neighborhoods (Davis, Gregory, and Hartley 2023), the insurance value of rent control (Diamond, McQuade, and Qian 2018), the willingness to pay to avoid violent crime and air pollution (Bishop and Murphy 2019), and horizontally differentiated consumption amenities (Almagro and Domínguez-Lino 2022), among others. We contribute to this literature by providing estimates on low-income renters' preferences over welfare-relevant neighborhood characteristics, levels of neighborhood attachment, and a rich set of moving costs by combining our detailed Census data with findings from the recent quasi-experimental shift-share literature.[9] We show how Black households appear to place less weight on access to employment opportunities relative to Non-Black households, but more weight on neighborhoods with higher shares of college graduates relative to Non-Black households. Our parameter estimates further show how the cost of moving across neighborhoods increases significantly with the length of households' residential tenure.

Third, our paper contributes to empirical research documenting the effects of gentrification on observable outcomes for low-income residents. Much of this research has focused on gentrification's impact on the propensity of incumbent residents to leave their home neighborhoods (Freeman and Braconi (2004), Freeman (2005), Ellen and O'Regan (2011), Ding, Hwang, and Divringi (2016), Dragan, Ellen, and Glied (2020)). This research has recently broadened to consider a wider range of outcomes. Baum-Snow, Hartley, and Lee (2019) examine the impact of neighborhood change on children's long run outcomes. Brummet and Reed (2021) consider effects on employment and experienced neighborhood characteristics alongside effects on residential mobility. Lester and Hartley (2014) and Meltzer and Ghorbani (2017) focus on the employment impacts of neighborhood change on incumbent residents. This literature broadly finds economically insignificant *average* effects on incumbents' household-level outcomes.[10] Statistical power, however, limits these studies' ability to examine heterogeneity across important dimensions like households' origin neighborhood environments. We advance this literature by showing that these average results are robust to incumbent households' origin neighborhood environments. This is a surprising finding as one might expect the impacts of gentrification to differ markedly based on local housing supply elasticities and the baseline level of neighborhood amenities.

[9]Our instrumental variable construction builds on Baum-Snow, Hartley, and Lee (2019) and Baum-Snow and Han (2023), who are the first to construct and microfound shocks to employment access in a model of workplace choice á la Tsivanidis (2022). Brummet and Reed (2021) and Glaeser, Luca, and Moszkowski (2023) use proximity to already gentrified tracts as an instrument for gentrification.

[10]Baum-Snow, Hartley, and Lee (2019) find meaningful impacts on incumbent children's future credit outcomes, consistent with the potential for neighborhood environments to effect children's outcomes in adulthood (Chyn and Katz 2021).

Roadmap. This paper is structured as follows. Section 2 introduces our data and our sample of low-income households. Section 3 defines our measure of gentrification and presents reduced-form evidence on the effects of gentrification on low-income incumbent renters. Section 4 details our dynamic model of neighborhood and workplace choice. Section 6 presents our parameters estimates and welfare comparisons. Section 7 concludes.

2. Data and Sample Construction

We use person- and establishment-level administrative micro data from the US Census Bureau spanning 2000-2019. Table 1 provides an overview of our data sources. We postpone a detailed exposition of the raw data and how we use it to construct our analysis samples to appendix C.1, presenting only a cursory discussion here.

TABLE 1. Data Description

Source	**Coverage**	**Description**
A. Household Panel		
Master Address File - Auxiliary Reference File (MAF-ARF)	2000-19	Annual Census address-level residential locations
Longitudinal Employer-Household Dynamics database (LEHD)	2000-19	Annual earnings, workplace locations, basic sociodemographics
American Community Survey (ACS)	2005-19	Detailed sociodemographics
B. Housing Characteristics		
CoreLogic	2006-2017	Address-level housing transactions and multiple listing service entries
Master Address File Extract (MAF-X)	2019	Address-level unit characteristics
C. Business Data		
Longitudinal Business Database (LBD)	2000-19	Establishment-level total employees and payroll

Data are obtained under ... The 2019 MAF-X is a continuously updated inventory of all known living quarters in the US. Addresses verified in the past, but that are no longer known living quarters, remain in the MAF-X except in rare circumstances. The MAF-X 2019 snapshot therefore contains an inventory of all known addresses spanning our entire sample period.

To construct our household panel, we start by forming an annual panel of persons' residential histories using the MAF-ARF. We merge to this residential history panel persons' annual earnings and workplace locations from the LEHD as well as additional sociodemographic characteristics from the ACS. These merges are facilitated by a unique person identifier called a protected identity key (PIK) which is assigned to individuals across data sets by the Census Bureau via probabilistic linking (Wagner and Layne 2014). We aggregate earnings by housing unit and designate the highest earner of each unit as the household head for that year.[11] Our household panel is then restricted to persons who have been identified as a household head and who occupy a rental housing unit.[12]

We restrict our sample to household heads that are between 25 and 65 years of age. To focus our analysis on low-income residents, we further restrict our sample to household heads earning in the bottom tercile of their respective core-based statistical area (CBSA) and decadal age band in the year they were first assigned household-head status.[13] We finally restrict our analysis to household heads residing in the urban cores of the 50 largest CBSAs that are present in the 28 states for which we can access data in the LEHD. We define urban cores using a similar method to Hwang and Lin (2016) and Couture and Handbury (2023), labeling them as the set of census tracts associated with each CBSA that contain the 50% of the CBSA's population that is closest to it's central business district (CBD).[14] Table 2 reports descriptive statistics for the sample of low-income renter households we use in our observational analysis.[15]

3. Observational Analysis

This section documents the impact of neighborhood change on several observed outcomes for incumbent renter households. We partition our outcomes into characteristics of households themselves (e.g. household income and commute distance), and characteristics of the neighborhoods that households reside in over time (e.g. neighborhood rents and socioeconomic composition). Before we share our results, we first define our measure of gentrification.

[11]We define housing units by their addresses in the MAF-X. Persons must have positive earnings in the given year to be considered as a household head.

[12]We describe how we impute rental unit status in appendix C.1. We also detail in appendix C.1 how we smooth residential histories and handle changes to household formations, out-of-sample migrations, and missing observations.

[13]CBSAs consist of counties associated with an urban core of at least 10,000 persons as well as adjacent counties that are deemed integrated through commuting ties.

[14]Our CBD definitions come from Fee and Hartley (2013). These CBD definitions are with respect to 2008 CBSA delineations. To keep consistent with these CBD definitions, we therefore use 2008 delineations of CBSAs throughout our analysis. Moreover, while each CBSA is associated with a primary urban center, some CBSAs additionally contain secondary urban centers called metropolitan divisions that have their own CBD. Although our low-income cutoff is constructed using the earnings distribution of the entire CBSA, our urban core cutoffs are particular to each urban center's population, including metropolitan divisions. We believe these choices best capture our target population of low-income urban residents.

[15]The sample we use to estimate our dynamic residential and workplace choice model is more expansive than the sample reported in Table 2, as it is not limited to household heads who are present in 2010 (the base year of our observational analysis). We are postponing the release of sample statistics for the full structural analysis panel to help streamline the US Census Bureau disclosure review processes.

Table 2. Sample Characteristics: Household Heads in 2002

	Black Households	Non-Black Households
Panel A: Household Head Characteristics		
Household Income	21,250	21,660
	(14,220)	(13,570)
Commute Distance	27.64	25.6
	(12.43)	(12.82)
College Degree	0.11	0.165
	(0.313)	(0.371)
Immigrant	0.212	0.45
	(0.409)	(0.498)
Age	42.38	43.08
	(10.64)	(10.8)
Female	0.611	0.519
	(0.488)	(0.5)
Household Size	2.544	2.505
	(1.549)	(1.578)
Parent	0.275	0.266
	(0.447)	(0.442)
Panel B: Household Heads' Tract Characteristics		
Median Rents	834.4	934.7
	(232.4)	(286.2)
Median Property Value	247,000	289,200
	(205,100)	(214,600)
Share White	0.435	0.692
	(0.264)	(0.208)
Share College-Educated	0.219	0.288
	(0.147)	(0.183)
Share College-Educated and White	0.124	0.213
	(0.137)	(0.169)
Distance to CBD	9.552	9.56
	(5.172)	(5.33)
Unique Households	314,000	688,000

Notes: Table reports mean characteristics with standard errors in parentheses. Sample consists of all persons in the panel designated as household heads in 2010. Panel A reports household head characteristics during 2010. Panel B reports characteristics of household heads' census tracts, also in 2010. Dollars are deflated to 2010 levels, and Census tracts are delineated by 2010 boundaries. Sources: 2005-2021 ACS, LEHD, CoreLogic, and MAF-ARF. Details on construction of tract aggregates are in the appendix. The Parent variable is calculated only for household heads present in the ACS 2005-2021, and is inferred by the reported age of the child in the year they are survey respondents. The college degree variable is only computed for PIKs for whose education variables are not imputed in the LEHD or for whom we ascertain educational attainment through our ACS surveys.

Defining Gentrification. We define gentrification as the increase in college-educated adults in neighborhood n from time t_0 to time t, normalized by the total adult population in n at time t_0:

$$\text{Gent}_{n,t_0,t} \equiv \frac{\text{College}_{n,t} - \text{College}_{n,t_0}}{\text{Adult Population}_{n,t_0}}$$

We follow the literature by defining gentrification in terms of educational attainment (as opposed to changes in racial shares or property values as often employed in the sociology literature), and by normalizing the change in college-educated adults by the total population in period t_0 (Brummet and Reed (2021); Card, Mas, and Rothstein (2008); Böhlmark and Willén (2020)). One alternative definition is the change in the *share* of college-educated households between periods t and t_0. We nonetheless follow the literature's convention since this measure minimizes the mechanical relationship between gentrification and a primary outcome of interest: the length of households' residential tenure in their origin neighborhood. This is because the residential choices of our sample throughout our analysis period has little influence on our measure of gentrification; only 11% and 16.5% of Black and non-Black households in our sample possess a college degree, and our sample of households comprises a small fraction of total households in each neighborhood.[16]

We choose to focus on the years 2010-2019 for our observational analysis. We do so for two primary reasons. First, this choice mitigates the influence of potentially confounding factors resulting from the Great Recession, especially as we can control for changing neighborhood-level characteristics prior to 2010. Second, restricting our panel to 2010-2019 allows us to control for household characteristics throughout 2000-2009 that are likely correlated with their residential location choices, such as their length of prior residential tenure.

3.1. Household Outcomes

We document no economically significant effects of neighborhood change on the income, the commute distance, or the duration of residential tenure of incumbent renter households. We first consider the effect of neighborhood change on the duration of time that incumbent residents remain in their origin neighborhood. We run Cox-Proportional Hazard models, estimating the impact that neighborhood change over 2010-2019 has on the probability an incumbent household leaves their origin neighborhood in any one year during this period (i.e. on incumbents' "hazard

[16] All the results we present below are quantitatively similar to specifications that exclude college-educated adults from our sample of low-income renters.

rates").[17] Our Cox-Proportional Hazards models take the following form:

$$\log(h(t|i)) = \alpha^{Cox} + \beta^{Cox}_{\mathrm{NC}}\mathrm{Gent}_{n(i)} + \gamma^{Cox}X_i + \delta^{Cox}X_{n(i)} + \alpha^{Cox}_{CBSA} + \varepsilon^{Cox}_i \tag{1}$$

where $h(t|i)$ is the hazard in period t for household i. $n(i)$ denotes this household's origin neighborhood, X_i is a vector of household-level controls, $X_{n(i)}$ is a vector of controls characterizing the origin neighborhood of household i, and α^{Cox}_{CBSA} is a CBSA-level fixed effect. We detail and motivate our choice of controls in Appendix D. To mitigate concerns over sample selection, we often restrict our sample to longtime renter households, defined as renter households who resided in their origin tract for at least five years prior to 2010. We also postpone discussion of this choice and identification more broadly to Appendix D. Throughout our reduced-form analyses, we cluster standard errors at the Census tract level, which is our treatment unit (Abadie et al. 2023).

Estimates of $\beta^{Cox}_{\mathrm{NC}}$ from equation 1 are reported in Panel A of Table 3. Columns (1) and (2) of Panel A in Table 3 report estimates for our full sample of renter householders, separately for Black and Non-Black headed households. Columns (3) and (4) report the same estimates but restrict to longtime renters. For all subsets of our data, we document economically insignificant effects of neighborhood change on incumbent renters' hazard rates. Consider the effect of neighborhood change on Non-Black longtime incumbent renters' hazard rate (column (4) in Panel A of Table 3). A 10 percentage point increase in our measure of neighborhood change corresponds to a 1.87 percent increase in the probability these renters leave their origin neighborhood in any given year between 2010 and 2019. Since the unconditional probability of leaving one's origin neighborhood in any one year reaches at most 20 percent, these effect sizes are negligible.[18] Table A3 documents that these null average results do not mask meaningful underlying heterogeneity with respect to the households' origin neighborhood environment. These results are consistent with the extant literature attributing neighborhood change to changes in in-migration patterns as opposed to increased exit rates among incumbent residents (Brummet and Reed 2021).

To explore the impact of gentrification on incumbent residents' future earnings, commute distances, as well as to test the robustness of our Cox-proportional hazards models, we estimate

[17] Our choice to estimate Cox-Proportional Hazard models is motivated by incumbent renters' short unconditional neighborhood tenures. Only 50.4 (51.7) percent of Black (non-Black) incumbent renter households remained in their 2010 origin census tract until at least 2015; these unconditional survival probabilities fall to 31.3 and 32.8 percent by 2019, respectively. Existing research that relies on intermittent sampling of residents therefore loses potentially identifying variation from incumbent residents with short unconditional residential tenures. The Cox-Proportional Hazard model allows us to efficiently utilize our full sample of residential histories to identify the effects of neighborhood change on incumbent renters' neighborhood tenures.

[18] A ten percentage point increase in gentrification increases the probability an incumbent renter household leaves their origin neighborhood by at most $(20/100) \times 1.87 = 0.374$ percentage points in any one year.

linear probability models of the form,

$$\Delta y_i = \alpha^{LP} + \beta_{\text{NC}}^{LP}\text{Gent}_{n(i)} + \gamma^{LP} X_i + \delta^{LP} X_{n(i)} + \alpha_{CBSA}^{LP} + \varepsilon_i^{LP} \tag{2}$$

where Δy_i denotes the change in a household-level outcome between 2010 and 2019. X_i, $X_{n(i)}$, and α_{CBSA}^{LP} are the same set of control variables and fixed effects used in equation 1. Estimates of β_{NC}^{LP} are reported in Panel B of Table 3. We again find economically insignificant effects of gentrification on our observed outcomes. For example, a 10 percentage point increase in our measure of gentrification increases annual earnings by only $287 in 2019 relative to 2010 for longtime Black renters. We find similarly economically insignificant estimates for changes in households' commute distances. The coefficients on the probability an incumbent household leaves their origin neighborhood during the analysis period are quantitatively consistent with our Cox proportional hazards model estimates.[19]

3.2. Experienced Tract Characteristics

In contrast to our household outcomes, we find economically significant impacts of gentrification on the neighborhood characteristics incumbent renters experience. This is especially true for Black renter households. A ten percentage point increase in our measure of gentrification in incumbent households' origin tracts led Black incumbent renters to reside in tracts during 2019 that had on average 2.7 percent higher rents and a 7.3 percent higher share of college educated adults. These average effects also mask meaningful heterogeneity across incumbent renters' origin neighborhood environments. The full set of results from our linear probability models in Appendix D report larger effect sizes for both changes in rents and college shares among longtime renters living in neighborhoods with an initially low share of college educated adults; incumbent Black (non-Black) households residing in neighborhoods with an initially low share of college-educated adults resided in neighborhoods with 9 percent (6 percent) higher rents during 2019 given a ten percentage point increase in our measure of gentrification. Similar patterns exist for changes in the share of college-educated adults across neighborhoods.

Our estimates on experienced neighborhood characteristics are indicative of moving frictions. For, without moving frictions, renters would simply re-optimize their location choice each period to ensure proximity to their ideal bundle of neighborhood characteristics, yielding economically insignificant estimates on experienced tract characteristics. Moving frictions in turn suggest the potential for differences in welfare effects from gentrification across tracts *within* CBSAs, motivating our paper's focus on *incumbent* renters. Moving frictions constrain

[19] Calculations showing this equivalency.

TABLE 3. Effect of Gentrification on Incumbent Renters

Outcome Variable	(1)	(2)	(3)	(4)
Panel A: Cox Proportional Hazards Model				
Hazard Rate	-0.0688	-0.107*	-0.0793	0.187*
	(0.0542)	(0.0439)	(0.132)	(0.0781)
Panel B: Linear Probability Model				
Leave Tract	-0.0176	-0.0263	-0.016	0.07*
	(0.0223)	(0.0174)	(0.0509)	(0.0278)
Income	2,872**	2,268**	-932	-86.05
	(948.8)	(736.4)	(2,170)	(1,375)
Commute Distance	-0.2280	-1.905***	0.637	0.5900
	(0.776)	(0.531)	(1.609)	(1.002)
Rent	0.271***	0.0960***	0.389***	0.222***
	(0.0472)	(0.0288)	(0.0620)	(0.0329)
College Share	0.726***	0.299***	1.079***	0.567***
	(0.204)	(0.0598)	(0.131)	(0.0788)
Sample Restrictions				
Race	Black	Non-Black	Black	Non-Black
Longtime Renters			✓	✓
N (1,000s)	314	688	56	156

Notes: Coefficients in Panel A correspond to the percent change in the hazard rate from a one hundred percentage point increase in our measure of gentrification. Coefficients in Panel B correspond to the impact of a one hundred percentage point increase in our measure of gentrification on a change in the associated outcome variable over 2010 to 2019. Leave Tract is an indicator equal to 1 if the household leaves their origin tract before 2019. Income is measured in 2010 dollars. Rent and college share are measured in percentage changes. Commute distances are measured in miles. Every specification includes the full set of controls listed and detailed in Appendix D. Standard errors in parentheses are clustered at the origin census tract level. Longtime renters are renters who have resided in their origin Census tract since at least 2005. These estimates were constructed using data from .... and were released under .... .

incumbent renters' choice sets, making them susceptible to local residential demand shocks among college-educated households. Whether incumbent renters then benefit from these demand shocks depends on their relative valuations of rents, job market access, and amenities vis-á-vis the actual change in these neighborhood characteristics.

It is worth noting here that there is limited correspondence between household residential mobility and welfare. Indeed, our estimates documenting no meaningful increase in neighborhood exit rates in response to gentrification are consistent with either positive or negative welfare effects from gentrification, depending on the aforementioned trade-off between rents, job market access, and amenities. Similarly, if we instead observed economically significant increases in neighborhood exit rates in response to neighborhood demand among college-educated households, we would need to understand whether these estimates reflected insignificant moving frictions and dense choice sets, or declines in incumbents' relative valuations of their origin neighborhoods (Vigdor 2002).

4. A Dynamic Model of Neighborhood and Workplace Choice

To quantify the welfare impact of neighborhood change on low-income incumbent renters, we estimate a dynamic discrete neighborhood and workplace choice model (Bayer et al. (2016); Diamond, McQuade, and Qian (2018); Davis et al. (2021); Davis, Gregory, and Hartley (2023); Almagro and Domínguez-Lino (2022)). We estimate our model to obtain parameter estimates of low-income households' preferences over neighborhood characteristics pertinent to understanding the welfare effects of gentrification. We use our estimated model parameters as well as observed changes in neighborhood rents, college shares, and employment access to compute the expected utility of incumbent renters separately for each low-income, urban census tract between 2000-2019. We finally difference these estimates with the expected utility renters would have obtained if the economy was in steady state during 2000, yielding tract-level estimates on the welfare effects of gentrification for incumbent renters. Note that while our expected utility calculations are constrained by our functional form assumptions, they never require us to compute counterfactual equilibria. We can therefore forego modeling the choices of agents beyond our sample population of low-income households.

Of papers in the dynamic discrete choice literature, our setup is most most similar to the neighborhood demand model in Almagro and Domínguez-Lino (2022), who analyze the endogenous formation of horizontally differentiated private consumption amenities in the context of Amsterdam's 2010-2019 tourism boom. One important departure from their demand model is that we incorporate differences in within-CBSA access to employment opportunities through a first-step workplace choice problem. We combine our LEHD and LBD data to compute microfounded, time-varying, and neighborhood-level measures of employment access for our sample population. To the extent that gentrification affects access to non-tradable employment opportunities, this addition is necessary to capture the full welfare effects from gentrification. Moreover, coupled with our job market access instrument described in section XX, these measures help facilitate credible identification of households' preferences over neighborhood characteristics.

4.1. Households and Timing of Choices

Each period, t, household heads, i, must decide which neighborhood in the city, c, they should live.[20] In addition to choosing their residential neighborhood, $n_{i,t}$, household heads must also decide which neighborhood to work in, $m_{i,t}$. Their workplace choice simply maximizes commute-time-discounted period income and is decided after neighborhood residence is known. Longer commute times reduce the time household heads spend working, effectively discounting the wage offered in workplace m. Finally, conditional upon deciding where to live and , households must then decide how much to spend on housing given the neighborhood-wide and period-specific rental rate.

Households differ ex-ante with respect to the household head's race, which we denote by $k \in$ {Black, Non-Black}. Households can further differ due to their previous neighborhood residences, which inform both their current neighborhood residence, $n_{i,t-1}$, as well as the how long they have lived there as of period $t-1$, $\tau_{i,t-1}$. We collect these observable household-level state variables in $x_{i,t} \equiv (n_{i,t-1}, \tau_{i,t-1})$.

Just before households make their workplace and neighborhood choices each period, they receive idiosyncratic productivity and preference shocks, respectively. These preference shocks are unobserved by the econometrician but rationalize observed variation in households' choices within types conditional on $n_{i,t-1}$ and $\tau_{i,t-1}$. The remainder of this section presents the household head's problem in reverse chronological order, starting with their workplace choice problem.

4.2. Workplace Choice

Upon making their residential neighborhood and housing consumption choices in period t, households receive two independent productivity shocks. The first productivity shock is denoted by b_t^{kc} and is common across all type-k households in city c. The second productivity shock is household- and workplace tract-specific. This second productivity shock is denoted by $z_{m,t}^i$, where m denotes workplace tract. Conditional on living in neighborhood n, households choose their work location to maximize their commute-time-discounted income:

$$I_{n,t}^k \equiv b_t^{kc} \cdot \max_m \frac{z_{m,t}^i}{d_{n,m}} w_{m,t},$$

where $w_{m,t}$ is the wage offered in workplace m and period t measured in efficiency units and $d_{n,m} > 0$ is time it takes to commute between neighborhood n and m. We assume households spend a fixed amount of time each day working or commuting, so $d_{n,m}$ effectively discounts the total wage offered in m: $z_{m,t}^i \cdot w_{m,t}$. We assume that $z_{m,t}^i$ is drawn independently from a

[20] Households take their city as given.

Frechet distribution with shape parameter $\epsilon^c \, \forall \, i \in c$. These shape parameters are specific to each city, which we make explicit with the superscript c. We further assume the Frechet shocks are independent across years, implying no cost to job switching. The expected income for a type-k household when living in tract n at time t is therefore,

$$\bar{I}^k_{n,t} = \Gamma\left(1 - \frac{1}{\epsilon^c}\right) \cdot b^{kc}_t \; \cdot \; RMA_n^{1/\epsilon^c},$$

where $\Gamma(\cdot)$ is the gamma function and $RMA_n \equiv \sum_{m \in \mathcal{N}^c} \left(\frac{w_m}{d_{n,m}}\right)^{\epsilon^c}$ is a summary measure of access to employment which we follow the literature in terming residential market access. We derive these equations and describe how we construct their empirical analogues in appendix B.3.

4.3. Neighborhood Choice

Households' Neighborhood Choice Problem. Households choose their residential locations to maximize the sum of their expected discounted utilities,

$$\max_{\{n \in \mathcal{N}^c\}_t^\infty} \quad \mathbb{E}\left[\sum_{t'=t}^{\infty} \delta^{t'-t} \cdot u^k_n\left(s^c_{i,t'}\right) \,\middle|\, \mathcal{I}_{i,t}\right] \tag{3}$$

where δ is a known discount factor, $s^c_{i,t'}$ is a vector of state variables that determine household i's flow utility u^k_n from choosing neighborhood n, and $\mathbb{E}\left[\cdot|\mathcal{I}_{i,t}\right]$ denotes the expectation operator conditioned on household i's information set at time t. $\mathcal{N}^c \equiv \{OO^c, 1^c, \ldots, N^c\}$ is the city-specific choice set, where OO^c denotes the outside option of leaving the city entirely. In each period, households observe the state variables s^c_{it} before choosing their residential location. Flow utilities are then realized and states evolve. Each household's information set $\mathcal{I}_{it}$ therefore includes all current and past state variables which they may use to form expectations over the evolution of the state variables. We are precise about households' beliefs in section 5.1.

State Variables. Households' flow utilities depend on the vector of state variables $s^c_{it} \equiv \left(x_{it}, \varepsilon_{int}, \omega^c_t, \xi^{kc}_t\right)$, where $\left(x_{it}, \varepsilon_{int}\right)$ are household-level observable and unobservable state variables, respectively. By contrast, $\left(\omega^c_t, \xi^{kc}_t\right)$ are city-specific observable and unobservable state variables, respectively. The household-level observable state variables, x_{it}, are comprised of households' residential tenure and neighborhood choice in the previous period, $x_{it} = (n_{it-1}, \tau_{it-1})$. The evolution of these observable household-level state variables are determined by household i's residential choices, as described in 4.1. ε_{int} is the household's unobservable state, which we assume is i.i.d. across households, neighborhoods, and time. We conceptualize ε_{int} as an unobserved-to-the-econometrician time-varying household and neighborhood-specific preference shock. As is common, we assume ε_{int} is distributed according to a Type I extreme value distribution.

Observable city-specific state variables are denoted by ω_t^c. The collection of city-specific state variables include vectors for each neighborhood's housing costs, r_{nt}^k, the share of college graduates in the neighborhood, $\frac{Coll_{nt}}{Pop_{nt}}$, each neighborhood's commute-time-discounted expected income, $\bar{I}_{nt}^k$, and an index for the time period t,[21]

$$\omega_t^c = \left(\{r_{nt}\}_{n\in\mathcal{N}^c}, \left\{ \frac{Coll_{nt}}{Pop_{nt}} \right\}_{n\in\mathcal{N}^c}, \left\{ \bar{I}_{nt}^k \right\}_{n\in\mathcal{N}^c}, t \right).$$

Expected commute-time-discounted income can vary across household types, hence the k superscript. Finally, ξ_t^{kc} is a city-specific vector of unobservable time-varying neighborhood-level amenity valuations among type k households. For example, ξ_t^{kc} could include time-varying valuations among type-k residents for suburban-life, independent of gentrification. To facilitate exposition, we define $\bar{\omega}_t^{kc} \equiv \left(\omega_t^c, \xi_t^{kc}\right)$ as the vector containing both observable and unobservable city-specific state variables.

Flow Utility. Preferences over neighborhood characteristics net of moving costs for a type-k household can be represented by

$$A_{n,t}^k \, Q_{n,t}^k \, \tau_{it}^{\beta_\tau^k} \, \exp(\varepsilon_{int})$$

where $A_{n,t}^k$ is a type-k's valuation of amenities in neighborhood n and $Q_{n,t}^k$ is consumption composite that is Cobb-Douglas over non-housing consumption, $C_{n,t}^k$, and housing consumption, $H_{n,t}^k$,

$$Q_{n,t}^k \equiv \left(C_{n,t}^k\right)^{\beta_C^k} \left(H_{n,t}^k\right)^{1-\beta_C^k}.$$

Households' expected period- and neighborhood-specific budget constraint is given by,

$$C_{n,t}^k \geq \bar{I}_{n,t}^k - r_{n,t} \cdot H_{n,t}^k.$$

Type-specific neighborhood amenities are,

$$A_{n,t}^k \equiv \left(\frac{Coll_{nt}}{Pop_{nt}} \right)^{\beta_A^k} \exp\left(\xi_{nt}^k\right).$$

[21] We include a time index in the set of observable neighborhood-level state variables to explicitly incorporate nonstationarity, so that the remaining observed state variables' evolutions can depend on time.

We can additionally decompose unobserved neighborhood- and period-specific amenities into time-invariant neighborhood-specific components, time-varying city-level components, as well as neighborhood-specific time varying components:

$$\xi_{nt}^{k} \equiv \alpha_{n}^{k} + \alpha_{t}^{kc} + \tilde{\xi}_{nt}^{k}.$$

Taking logs, incorporating moving costs (defined below in 4.3), solving for expected optimal housing consumption, and substituting in the amenity specification yields the following expected flow-utility specification for a type-k household choosing neighborhood n with state s_{it}^{c} that is consistent with households' preferences over neighborhood characteristics,[22]

$$\begin{aligned} u_{n}^{k}\left(s_{it}^{c}\right) = \alpha_{n}^{k} + \alpha_{t}^{kc} + \beta_{w}^{k} \ln\left(\bar{I}_{n,t}\right) - \beta_{r}^{k} \log\left(r_{n,t}\right) + \beta_{A}^{k} \ln\left(\frac{Coll_{nt}}{Pop_{nt}}\right) \\ + \beta_{\tau}^{k} \ln(\tau_{it}) - MC_{t}^{k}(n_{t}, n_{it-1}) + \tilde{\xi}_{nt}^{k} + \varepsilon_{int} \end{aligned}$$

Moving Costs. If a type-k household decides to leave their current neighborhood for another neighborhood in the same city they incur a non-monetary moving cost, $MC_{t}^{k}(n_{it}, n_{it-1})$, that is composed of a fixed disutility from moving, the physical straight-line distance between the household's origin and destination neighborhoods, as well as the social distance between these two neighborhoods. Conversely, if a type-k household decides to leave their city entirely, they obtain a single city-specific fixed cost. Specifically,

$$MC_{t}^{k}(n_{t}, n_{t-1}) = \begin{cases} 0 & \text{if } n_{t} = n_{t-1} \\ MC^{k} + \boldsymbol{\beta}_{d}^{k\prime} \boldsymbol{d}(n_{t}, n_{t-1}) + \boldsymbol{\beta}_{s}^{k\prime} \boldsymbol{s}(n_{t}, n_{t-1}) & \text{if } n_{t} \neq n_{t-1} \text{ and } n_{t}, n_{t-1} \neq OO^{c} \\ MC^{kc} & \text{if } n_{it} \neq n_{t-1} \text{ and } n_{t} \text{ or } n_{t-1} = OO^{c}. \end{cases}$$

where MC^{k} and MC^{kc} are the fixed intensive and extensive moving costs, respectively. $\boldsymbol{d}(n_{t}, n_{t-1})$ is a vector describing the distance between n_{t}, n_{t-1} and $\boldsymbol{s}(n_{t}, n_{t-1})$ is a vector describing the social distance between n_{t}, n_{t-1} in period t,

$$\boldsymbol{d}(n_{t}, n_{t-1}) \equiv \begin{bmatrix} |Dist(n_{t}, n_{t-1})| \\ |Dist(n_{t}, n_{t-1})|^{2} \end{bmatrix} \quad \boldsymbol{s}_{t}(n_{t}, n_{t-1}) \equiv \begin{bmatrix} |(S(n_{t}) - S(n_{t-1}))/(S(n_{t}) + S(n_{t-1})| \\ |(S(n_{t}) - S(n_{t-1}))/((S(n_{t}) + S(n_{t-1}))|^{2} \end{bmatrix}$$

where $Dist(n_{t}, n_{t-1})$ is the straight-line distance between the centroids of neighborhood n_{t} and n_{t-1}, and $S(n_{t})$ is the share of college graduates in neighborhood n in time t.[23] Our measure of

[22]Flow utility is expected in that it is the value households expect to obtain before their period-specific workplace tract productivity shocks are realized. It is this flow utility specification that is relevant for households' neighborhood choice and thus for our structural estimation.

[23]We normalize our social distance measure by the sum of shares so the percentage changes are invariant to the direction of the residential move.

social distance captures the fact that while low-income renter households may value residing in neighborhoods with a high share of college graduates, it may be costly to assimilate to neighborhood environments different from one's own (Gans (1982), Jargowsky (2009)). Indeed, recent experimental research suggest low-income households' moving costs are poorly approximated by the physical distance of residents' potential moves, but strongly predicted by differences in the sociodemographic composition of households' origin and potential destination neighborhoods (Bergman et al. 2023).

Value Functions, Choice Probabilities, and Expectational Errors. We denote $V^k(s_{it}^c)$ as the value function of the dynamic programming problem associated with equation 3. By Bellman's principle of optimality,[24]

$$V^k\left(s_{it}^c\right) = \max_{n\in\{OO^c,1^c,\ldots,N^c\}} \left\{\mathbb{E}_{x'|x,n}\left[u_n^k\left(s_{it}^c\right)\right] + \delta\mathbb{E}_t\left[V^k\left(s_{it+1}^c\right)|n, s_{it}^c\right]\right\}$$

We define household i's ex-ante continuation value function as the expectation of the value function with respect to ε_{int},

$$\bar{V}^k\left(x_{it}, \bar{\omega}_t^{kc}\right) \equiv \int V^k\left(s_{it}^c\right) dF^{\varepsilon}(\varepsilon_{int}), \tag{4}$$

and household i's conditional value function as

$$\begin{aligned} v_n^k\left(x_{it}, \bar{\omega}_t^{kc}\right) &\equiv \mathbb{E}_{x'|x,n}\left[u_n^k\left(s_{it}^c\right)\right] - \varepsilon_{int} + \delta\mathbb{E}_t\left[\bar{V}^k\left(x_{it+1}, \bar{\omega}_{t+1}^{kc}\right)|n, x_{it}, \bar{\omega}_t^{kc}\right] \\ &\equiv \bar{u}_n^k\left(x_{it}, \bar{\omega}_t^{kc}\right) + \delta\mathbb{E}_t\left[\bar{V}^k\left(x_{it+1}, \bar{\omega}_{t+1}^{kc}\right)|n, x_{it}, \bar{\omega}_t^{kc}\right]. \end{aligned} \tag{5}$$

Then, given our assumption that ε_{int} are distributed i.i.d Type I extreme value, the probability a type-k household with state variables $\left(x_{it}, \bar{\omega}_t^{kc}\right)$ chooses neighborhood n in period t is given by,

$$p_n^k\left(x_{it}, \bar{\omega}_t^{kc}\right) = \frac{\exp\left(v_n^k\left(x_{it}, \bar{\omega}_t^{kc}\right)\right)}{\sum\limits_{n'\in\mathcal{N}^c}\exp\left(v_{n'}^k\left(x_{it}, \bar{\omega}_t^{kc}\right)\right)}, \tag{6}$$

and the ex-ante value function in 4 has the value

$$\bar{V}^k\left(x_{it}, \bar{\omega}_t^{kc}\right) = \ln\left(\sum_{n\in\mathcal{N}^c}\exp\left(v_n^k\left(x_{it}, \bar{\omega}_t^{kc}\right)\right)\right) + \gamma,$$

[24] The expectation operator $\mathbb{E}_{x'|x,n}[\cdot]$ is with respect to the future value of households' observed household-level state variables, x', conditional on households' current state and on their neighborhood choice. While the current deterministic setup yields this operator redundant, we include it here to be consistent with our empirical application that models the evolution of households' residential tenure stochastically conditional on their neighborhood choice.

where γ is Euler's constant. Combining these two expressions yields the following well-known result which is critical to deriving our estimating equations (Hotz and Miller 1993),

$$\bar{V}^k\left(x_{it}, \bar{\omega}_t^{kc}\right) = v_n^k\left(x_{it}, \bar{\omega}_t^{kc}\right) - \ln\left(p_n^k\left(x_{it}, \bar{\omega}_t^{kc}\right)\right) + \gamma. \tag{7}$$

Another critical element for deriving our estimating equations is the difference between households' expected ex-ante continuation values and their realized counterparts,

$$e^{\bar{V}}(x', \bar{\omega}_t^{kc}, \bar{\omega}_{t+1}^{kc}) \equiv \underbrace{\bar{V}(x', \bar{\omega}_{t+1}^{kc})}_{\text{realized}} - \mathbb{E}_{\bar{\omega}'|\bar{\omega}_t^{kc}}\left[\bar{V}(x', \bar{\omega}')\middle|\bar{\omega}_t^{kc}\right]. \tag{8}$$

We follow Kalouptsidi, Scott, and Souza-Rodrigues (2021) and term the differences *expectational errors*. These expectational errors allow us to discard households' actual expectations in estimation. Having to solve for households' expectations would be prohibitively costly given the high-dimensional nature of a household's state-space (some urban cores have over a thousand 2010-delineated census tracts).

Now that the function dependencies are clear, going forward we suppress their arguments and remove the city superscripts unless they are required for explicative purposes: $\bar{V}_{xnt}^k \equiv \bar{V}^k\left(x_{it}, \bar{\omega}_t^{kc}\right)$, $V_{int}^k \equiv V^k(s_{it}^c)$, $\bar{u}_{xnt}^k \equiv \bar{u}_n^k\left(x_{it}, \bar{\omega}_t^{kc}\right)$, $v_{xnt}^k \equiv v_n^k\left(x_{it}, \bar{\omega}_t^{kc}\right)$, and $p_{xnt}^k \equiv p_n^k\left(x_{it}, \bar{\omega}_t^{kc}\right)$.

5. Structural Estimation

We estimate our neighborhood demand parameters with what is termed in the dynamic discrete choice literature an Euler Equations in Conditional Choice Probabilities (ECCP) estimator.[25] Like other conditional choice probability estimators, ECCP estimation involves two steps. In the first step, we estimate households' conditional choice probabilities, the soon-to-be-introduced transition distributions for the household-level state variables, as well as households' variable moving costs $\boldsymbol{\beta}_d^{k\prime}$ and $\boldsymbol{\beta}_s^{k\prime}$. We then estimate the remaining model parameters in a second step, conditional on our first-step estimates. Estimation in the second step is based off moment restrictions implied by the dynamic optimization of households, which we derive and explain in section 5.2. With flow utilities linear in the model's parameters, we can evaluate these moment restrictions in a standard linear GMM framework.[26]

[25]ECCP estimators are termed as such since they can be viewed as discrete choice analogues to Euler equations in models with continuous choice variables (Aguirregabiria and Magesan (2013), Kalouptsidi, Scott, and Souza-Rodrigues (2021)).

[26]ECCP estimation has been used in a variety of applied settings, from choices over agricultural land use, (Scott 2013), new technology adoption, (Groote and Verboven 2019), occupational choice (Traiberman (2019), Gendron-Carrier (2023)), and most relevant to our setting, residential neighborhoods (Diamond, McQuade, and Qian (2018), Davis et al. (2021), Almagro and Domínguez-Lino (2022)). See Kalouptsidi, Scott, and Souza-Rodrigues (2021) for a

The ECCP estimator has many advantages in our setting. First, the ECCP estimator is computationally light. Since our analysis covers the residential history of low-income households for ten years in 50 large US metropolitan areas at the Census tract level, traditional dynamic discrete choice estimation procedures that explicitly solve for households' value functions are infeasible (e.g. Rust (1987)). Second, our focus on gentrification implies an inherently nonstationary environment, making modeling the evolution of neighborhood change conceptually and computationally challenging. As we demonstrate in the derivation of our moment restrictions, ECCP estimation does not require a complete specification of households' information sets and therefore neither the evolution of our city-specific state variables. Third, by providing moment conditions that we can evaluate with standard linear regression techniques, we can exploit our instrumental variables detailed in section 6 to estimate our model parameters in a manner that is consistent with the recent literature on identification using Bartik shift-share instruments (Goldsmith-Pinkham, Sorkin, and Swift (2020), Borusyak, Hull, and Jaravel (2022)).

5.1. Estimation Assumptions

To identify our neighborhood demand parameters, we must make the following set of assumptions:

(a) *State Transitions*: The state variables s^c_{it} evolve according to a controlled first-order Markov process with a transition distribution that factors as,[27]

$$f\left(s^c_{it+1}|n_{it}, s^c_{it}\right) = f^x(x_{it+1}|n_{it}, x_{it}) \cdot f^{\bar{\omega}}\left(\bar{\omega}^{kc}_{t+1}|\bar{\omega}^{kc}_t\right) \cdot f^{\varepsilon}(\varepsilon_{int+1}).$$

(b) *Utility Normalization*: The utility offered by the outside option in every city is normalized to α^{ck} for each time period,

$$\alpha^k_{OO^c} + \beta^k_w \ln(\bar{I}_{OO^c,t}) - \beta^k_r \ln(r^k_{OO^c,t}) + \beta^k_A \ln\left(\frac{Coll_{nt}}{Pop_{nt}}\right) + \tilde{\xi}^k_{OO^c t} = \alpha^{ck} \quad \forall\, t.$$

(c) *Rational Expectations*: Households' expectations over the evolution of the CBSA-level state variables conditional on their information set $\mathcal{I}_{it}$ correspond to the conditional expectations

comprehensive econometric treatment of linear regression techniques with ECCP estimators.

[27] The evolution of the individual-level state variables, x_{it}, are "controlled" in that their evolution is influenced by the household's choices. While the current setup ensures the evolution of x_{it} is fully determined by the household's neighborhood choice, our empirical implementation assumes τ_{it} evolves stochastically conditional on the household's choice in order to reduce the dimensionality of our problem. Whether τ_{it} evolves stochastically or not, our discussion surrounding households' choice sets and moving costs in section 4.3 should make clear that x_{it+1} evolves independently of $\bar{\omega}^{kc}_t$ conditional on n_{it} and x_{it}.

of the true data generating process given $\mathcal{I}_{it}$,

$$\mathbb{E}\left[e^{\bar{V}}(x', \bar{\omega}_t^{kc}, \bar{\omega}_{t+1}^{kc})|\mathcal{I}_{it}\right] = 0,$$

where $e^{\bar{V}}(x', \bar{\omega}_t^{kc}, \bar{\omega}_{t+1}^{kc})$ are the expectational errors were defined in equation 8.

An important implication of assumption (a) is that the market-level state variables $\bar{\omega}_t^{kc}$ are *perceived* as exogenous by individual households; a household cannot expect to individually affect the evolution of $\bar{\omega}_t^{kc}$ with their own neighborhood choice (cf. assumption (1) in Kalouptsidi, Scott, and Souza-Rodrigues (2021)). Given the typical 2010 US census tract houses around 4,000 residents, we believe this assumption is plausible.[28] Note that assumption (a) does not require the observed and unobserved city-specific state variables to evolve independently of one another. We highlight this to foreshadow the econometric challenge we face when attempting to identify preferences over functions of ω_t^c.

Assumption (b) says that residents who choose to reside outside of their respective CBSA's urban core obtain a time-invariant and CBSA-specific mean utility. [29] This assumption normalizes each CBSA's neighborhood mean utilities to a constant and time-invariant level, which is necessary to compare welfare across households within-CBSAs given logit models only identify differences in mean utilities. The assumption moreover facilitates exposition and, because each α^{ck} is unobserved, highlights the incommensurability of expected welfare both across different k-types and across CBSAs.

Lastly, assumption (c) says that, on average, households correctly anticipate the evolution of $\bar{\omega}_t^{kc}$. An important corollary of assumption (c) is that the contents of households' information sets in time t are mean independent of their expectational errors also at time t. The importance of this corollary will become clear in section 5.3 when we discuss choosing among our set of instrumental variables to estimate households' preferences.

5.2. Deriving our Estimating Equations

Our goal now is to take our model setup and show how one can derive estimating equations linear in households' demand parameters. With estimating equations linear in households' demand parameters, we may estimate our model with standard linear GMM techniques. To derive these equations, though, we must first introduce the concept of renewal actions.

[28]While our estimating equations are implied by households' dynamic optimization, they are not informed by any equilibrium conditions, allowing us to remain agnostic over how households' individual choices influence the evolution of the CBSA-wide state variables.

[29]Recall that *all* residents of a given CBSA (i.e. including those outside the urban core) additionally receive a time-varying but neighborhood-invariant utility shock, α_t^k. The value of the outside option can therefore shift over time, but always in proportion to the mean utilities of inner-city neighborhoods.

Renewal Actions. To derive our estimating equations, we make use of what are termed in the dynamic discrete choice literature "renewal actions" (Hotz and Miller (1993), Arcidiacono and Miller (2011)). Renewal actions are actions that when taken in period t lead to the same distribution of states at the beginning of period $t+1$, regardless of the household's state in period t. In our setting, simply moving to a new neighborhood is a renewal action; moving to a new neighborhood resets a household's residential tenure to 0 regardless of the household's origin neighborhood or their current residential tenure. Moreover, because the city-specific state variables, $\bar{\omega}_{t+1}^{kc}$, and unobserved idiosyncratic preference shocks, ε_{int}, are independent of the household's state in period t, *all* the remaining state variables are reset to a common value upon moving to a new neighborhood.[30]

We exploit such renewal actions when deriving our estimating equation. To see how, consider the residential choices of two hypothetical type-k households between periods $t-1$ and $t+1$. Assume that in period $t-1$ these households reside in the same neighborhood, n_{t-1}, but not in period t (i.e. at least one household chooses a new neighborhood in period t). Further, assume that in period $t+1$ both households move to the same neighborhood, n_{t+1}. Our estimation procedure involves relating the difference in the expected discounted utilities of the two households' neighborhood choices to the difference in the probability that these neighborhood choices are actually made. Critically, because moving to neighborhood n_{t+1} in period $t+1$ constitutes the same renewal action for both households, their state variables are reset to a common value which in turn equalizes their continuation values. Differences in the expected discounted utilities associated with the two sets of neighborhood choices are thus only a function of households' flow utilities. Relating such differences in expected discounted utilities to households' choice probabilities in this way therefore helps disentangle observed variation in households' flow utilities from households' unobserved continuation values.

To ease exposition, moving forward we term these consecutive residential location choices *residential paths*.

Our Estimating Equation. Consider the set of residential paths we just described for type-k households but with an additional requirement that one of the households chooses the outside option in period t:

(a) In period $t-1$, both households reside in the same neighborhood, n_{t-1}.

(b) In period t, one household chooses neighborhood n, while the other household chooses n'.

[30] Note that renewal actions depend on our construction of households' neighborhood tenure in section 4.3. This construction assumes that the length of households' prior residential tenures has no impact on the value to future residential tenure. While this is a strong assumption, it is necessary to keep the dimension of the household-level state-space manageable.

n' also happens to be the outside option. While it must be the case that $n' \neq n$, it may be that $n_{t-1} = n$ or $n_{t-1} = n'$.

(c) In period $t+1$, both households convene at $\tilde{n}$, where $\tilde{n} \neq n'$ and $\tilde{n} \neq n$.

Given these residential paths, we can derive equation 9 which is linear in households' preference parameters. It is this equation that we use to construct moment conditions that identify households' preferences. We derive it by equating the difference in the expected discounted utilities associated with the two residential paths to the difference in the probability that households actually take these paths,

$$Y^k_{xnn'\tilde{n}t} = \tilde{\alpha}^k_n + \alpha^k_t + \beta^k_w \ln(\bar{I}_{OO^c,t}) - \beta^k_r \ln(r^k_{OO^c,t}) + \beta^k_A \ln\left(\frac{Coll_{nt}}{Pop_{nt}}\right) + \beta^k_\tau \tilde{\tau}_x - \widetilde{MC}^k_t + \nu^k_{xnn'\tilde{n}t}, \tag{9}$$

where

$$\begin{aligned}
Y^k_{xnn'\tilde{n}t} &\equiv \ln\left(\frac{p^k_{xnt}}{p^k_{xn't}}\right) - \delta\left(\sum_{x'} \ln\left(p^k_{x\tilde{n}t}\right) f^x\left(x'|n, x_t, \bar{\omega}^{kc}_t\right) - \sum_{x'} \ln\left(p^k_{x\tilde{n}t}\right) f^x\left(x'|n', x_t, \bar{\omega}^{kc}_t\right)\right)\\
\tilde{\alpha}^k_n &\equiv \alpha^k_n - \alpha^{ck}\\
\tilde{\tau}_x &\equiv \sum_{x'} \ln(\tau_{xt}(x')) f^x\left(x'|n, x_t, \bar{\omega}^{kc}_t\right) - \sum_{x'} \ln(\tau_{xt}(x')) f^x\left(x'|n', x_t, \bar{\omega}^{kc}_t\right)\\
\widetilde{MC}^k_t &\equiv MC^k_t(n, n_{xt-1}) - MC^k_t(n', n_{xt-1}) - \delta\left(MC^k_t(\tilde{n}, n) - MC^k_t(\tilde{n}, n')\right)\\
\nu^k_{xnn'\tilde{n}t} &\equiv \tilde{e}(x_t, \bar{\omega}^{kc}_t, \bar{\omega}^{kc}_{t+1}) + \tilde{\xi}^k_{nt}.
\end{aligned}$$

Since the full derivation of this estimating equation is becoming well known, we relegate it to appendix B.1. With estimates of households' conditional choice probabilities, $\hat{p}^k$, and estimates of the household-level transition distributions, $\hat{f}^x$, we can estimate equation 9 using linear GMM. The next subsection details our two-step estimation procedure.

5.3. Two-Step Estimation Procedure

In the first step of the estimation procedure we estimate i) transition probabilities for the household-level state variables; ii) households' conditional choice probabilities; and iii) households' variable moving cost parameters. With these estimates in hand, we can estimate equation 9 using linear GMM.

Household Transition Distributions. To keep the dimension of the household state-space manageable, we follow the literature stemming from Rust (1987) and discretize our household-level residential tenure measure into two buckets.[31] Specifically, we aggregate tenure into two buckets,

$$\bar{\tau} = \begin{cases} 1 & \text{if } \tau \leq 4 \\ 2 & \text{otherwise.} \end{cases}$$

We assume that this aggregated location tenure variable evolves stochastically according to the following distribution function:[32]

$$f^{\bar{\tau}}(\bar{\tau}_t = 1 | n_t, x(n_{t-1}, \bar{\tau}_{t-1}), \bar{\omega}^{kc}) = 1 \text{ if } n_t \neq n_{t-1}$$
$$f^{\bar{\tau}}(\bar{\tau}_t = 2 | n_t, x(n_{t-1}, \bar{\tau}_{t-1}), \bar{\omega}^{kc}) = \begin{cases} 1, & \text{if } n_t = n_{t-1} \text{ and } \bar{\tau}_{t-1} = 2 \\ g^k_{n_{t-1}}, & \text{if } n_t = n_{t-1}, \bar{\tau}_{t-1} = 1, \text{ and } i \in k. \end{cases}$$

where we estimate $g^k_{n_{t-1}}$ directly from the data,

$$\hat{g}^k_{n_{t-1}} = \frac{\sum\limits_{i \in n_t, k} \mathbb{1}\{\tau_{xt} = 5\}}{\sum\limits_{i \in n_{t-1}, k} \mathbb{1}\{\tau_{xt-1} \leq 4\}}.$$

Given that our analysis is at the Census tract level, estimated in this way $\hat{g}^k_{n_{t-1}}$ is a sparse measure of $g^k_{n_{t-1}}$. In practice we therefore take a weighted average of $\hat{g}^k_{n_{t-1}}$ across census tracts in each county and each year.

Conditional Choice Probabilities. Researchers typically face a trade-off between sparsity and flexibility when estimating first-step conditional choice probabilities, $\hat{p}^k_{xnt}$. Our setting is no different. On the one hand, we may estimate $\hat{p}^k_{xnt}$ directly from the data by calculating the probability a type-k household with state x_{it} moves to each neighborhood $n_t \in \mathcal{N}^c$.[33] While this approach does not impose any restrictions on the implied data generating process, it leads to very sparse estimates of $\hat{p}^k_{xnt}$ in our setting given the number of census tracts in our largest CBSAs. On the other hand, we may impose some structure on the implied data generating process to smooth $\hat{p}^k_{xnt}$. Given our particularly sparse empirical choice probabilities, we choose this latter option.

We model the count of type-k households in neighborhood $n \in \mathcal{N}^c$ choosing neighborhood $n' \in \mathcal{N}^c$ between periods t and $t+1$ as being derived from a Poisson distribution.[34] We parame-

[31]The remaining household-level state variable is the household's residential location in the previous year. This state variable evolves deterministically depending upon the residential path under consideration. We therefore do not need to specify any transition probability for this component of x_t.

[32]Our notation here corresponds to the marginal transition distribution of τ_{it}, taking n_{it} as given.

[33]We could similarly employ any non-parameteric methods to compute $\hat{p}^k_{xnt}$ directly from the data.

[34]We choose to model the data as a Poisson distribution because of its ability to account for sparse data and its

terize the mean of the Poisson distribution in a way that does not impose additional restrictions on the data generating process implied by our dynamic model.[35] Specifically, we estimate the following flexible Poisson regression separately for each type-k household,

$$\log\left(\mathbb{E}\left[\left|n_{t-1}^{k\bar{\tau}} \to n_t^{\prime k\bar{\tau}}\right|\right]\right) = \gamma_{n't}^k + \mu_{\bar{\tau}} \cdot \mathbb{1}\{n' = n_{t-1})\} + \gamma_{n't}^k \cdot \lambda_{\bar{\tau}} \cdot \mathbb{1}\{n' = n_{t-1})\} - MC_t^k(n', n_{t-1}), \tag{10}$$

where $\left|n_{t-1}^{k\bar{\tau}} \to n_t^{\prime k\bar{\tau}}\right|$ is the count of type-k households with aggregated tenure status $\bar{\tau}$ in neighborhood n that choose neighborhood n' between period $t-1$ and t. $\gamma_{n't}^k$ is a fixed effect that captures the neighborhood- and period-specific component of utility associated with choosing neighborhood n' in period t, $\mu_{\bar{\tau}}$ is a fixed effect that captures the additional utility residents obtain from staying in their origin neighborhood given their tenure status, $\bar{\tau}$, and $\lambda_{\bar{\tau}}$ is a parameter capturing how the additional value one obtains from staying in their origin neighborhood varies with neighborhood mean utilities. $\mathbb{1}\{n' = n_{t-1})\}$ is an indicator variable that equals 1 if the household stays in their origin tract, and $MC_t^k(n', n_{t-1})$ are the same moving costs described in section 4.3. We use our estimates from the Poisson model to predict the probability a type-k household with aggregated tenure status $\bar{\tau}$ living in neighborhood n chooses neighborhood n' in each year: $\hat{p}_{xn't}^k$.[36] As expected, the predicted probabilities are strongly correlated with their empirical counterparts; the coefficients of correlation are 0.951 and 0.984 for Black and non-Black households, respectively.

Equation 10 additionally identifies our variable moving cost parameters, $\boldsymbol{\beta}_d^{k\prime}$ and $\boldsymbol{\beta}_s^{k\prime}$. Since the cost of moving to neighborhood n differs for each type-k household depending on their origin neighborhood, we can separately identify the parameters governing variable moving costs from γ_{nt}^k, $\mu_{\bar{\tau}}$, and $\lambda_{\bar{\tau}}$. How the cost of moving varies with the physical distance of the move, $\boldsymbol{\beta}_d^{k\prime}$, is identified with variation in the distance households move within their urban core, conditional on moving. Similar variation but with respect to social distance identifies the cost of moving to neighborhoods socially different to one's origin neighborhood, $\boldsymbol{\beta}_s^{k\prime}$.[37] We report our variable moving cost estimates with the rest of our parameter estimates in table 4.

computational efficiency (Correia, Guimarães, and Zylkin 2020).

[35]Note that the independence of households' neighborhood moves in any one period implied by the Poisson distribution is embedded in assumption (a); f^x for household i is independent of all other households' actions. Appendix B.2 additionally shows how our Poisson regression specification does not imply additional restrictions on the neighborhood choice problem of households in our dynamic model.

[36]Many Census tracts lack type-k households or in-migrants in a given year. Consequently, we cannot compute their conditional choice probabilities and exclude them from type-k households' choice sets. These tracts tend to be suburban or affluent, and their removal is unlikely to affect our estimates.

[37]Note that in this repeated cross-sectional framework, we are unable to separately identify the fixed cost of moving from the value of residential tenure. We must instead estimate households' fixed moving costs in a second step using equation 9. Note also that our assumption of a Poisson distribution in equation 10 does not impact the variable moving cost estimates we obtain here. This is because of the isomorphism between the score of the Poisson distribution and of the conditional logit for these continuous variables, yielding identical MLE estimates (Guimarães, Figueirdo, and Woodward 2003).

Step Two. Given our estimated variable moving cost parameters, our estimated conditional choice probabilities, and our estimated transition probabilities from step 1, we may now construct the empirical analogue of equation 9:

$$\hat{Y}^k_{xnn'\tilde{n}t} = \tilde{\alpha}^k_n + \alpha^k_t + \beta^k_w \ln(\bar{I}_{n,t}) - \beta^k_r \ln(r_{n,t}) + \beta^k_A \ln\left(\frac{Coll_{nt}}{Pop_{nt}}\right) + \beta^k_\tau \hat{\tilde{\tau}}_{xt} - \widetilde{\widehat{MC^F}}^k_t + \nu^k_{xnn'\tilde{n}t}, \tag{11}$$

where

$$\begin{aligned}
\hat{Y}^k_{xnn'\tilde{n}t} &\equiv \ln\left(\frac{\hat{p}^k_{xnt}}{\hat{p}^k_{xn't}}\right) - \delta\Bigg(\sum_{x'} \ln\left(\hat{p}^k_{x\tilde{n}t}\right) \hat{f}^x\left(x'|n, x_t, \bar{\omega}^{kc}_t\right) \\
&\qquad - \sum_{x'} \ln\left(\hat{p}^k_{x\tilde{n}t}\right) \hat{f}^x\left(x'|n', x_t, \bar{\omega}^{kc}_t\right)\Bigg) + \widetilde{\widehat{MC^V}}^k_t \\
\hat{\tilde{\tau}}_{xt} &\equiv \sum_{x'} \ln(\bar{\tau}_{xt}(x'))\hat{f}^x\left(x'|n, x_t, \bar{\omega}^{kc}_t\right) - \sum_{x'} \ln(\bar{\tau}_{xt}(x'))\hat{f}^x\left(x'|n', x_t, \bar{\omega}^{kc}_t\right) \\
\widetilde{\widehat{MC^\upsilon}}^k_t &\equiv \widehat{MC^\upsilon}^k_t(n, n_{t-1}) - \widehat{MC^\upsilon}^k_t(n', n_{t-1}) - \delta\left(\widehat{MC^\upsilon}^k_t(\tilde{n}, n) - \widehat{MC^\upsilon}^k_t(\tilde{n}, n')\right) \\
\nu^k_{xnn'\tilde{n}t} &\equiv \tilde{e}(x_t, \bar{\omega}^{kc}_t, \bar{\omega}^{kc}_{t+1}) + \tilde{\xi}^k_{nt},
\end{aligned}$$

and where $\hat{\cdot}$ represents estimates from the first step. $\widetilde{\widehat{MC^\upsilon}}^k_t$ is the difference in the fixed, $\upsilon = F$, or variable, $\upsilon = V$, portion of moving costs. We detail the difference in expectational errors, $\tilde{e}(x_t, \bar{\omega}^{kc}_t, \bar{\omega}^{kc}_{t+1})$, in appendix B.1.

To be precise about identification, it is worth unpacking the error term, $\nu^k_{xnn'\tilde{n}t}$, in equation 11. $\nu^k_{xnn'\tilde{n}t}$ is comprised of both unobserved neighborhood-specific amenities, $\tilde{\xi}^{kc}_{nt}$, as well as expectational errors, $\tilde{e}(x_t, \bar{\omega}^{kc}_t, \bar{\omega}^{kc}_{t+1})$. We will consider each of these in turn starting with the unobserved neighborhood-specific amenities, $\tilde{\xi}^{kc}_{nt}$. Since we place no restriction on the relationship between $\tilde{\xi}^{kc}_{nt}$ and the remaining time-varying observable neighborhood characteristics, OLS estimates of 11 will be biased. Expected income, neighborhood-level housing costs, and the share of college graduates will invariably be correlated with many unobserved neighborhood-level factors like proximity to natural amenities that we do not observe as econometricians.

To distinguish between preferences for observed versus unobserved neighborhood amenities, we start by differencing equation 11 using residential paths starting in 2017 (i.e. $t-1 = 2017$) and residential paths starting in 2010 (i.e. $t-1 = 2010$), obtaining,

$$\Delta\hat{Y}^k_{xnn'\tilde{n}} = \Delta\alpha^k_t + \beta^k_w \Delta\ln(\bar{I}_{n,t}) - \beta^k_r \Delta\ln(r_{n,t}) + \beta^k_A \Delta\ln\left(\frac{Coll_n}{Pop_n}\right) + \Delta\nu^k_{xnn'\tilde{n}}, \tag{12}$$

where the Δ's correspond to the difference in the associated variables between $t = 2011$ and

$t = 2018$. Differencing equation 11 removes the time-invariant component of exogenous neighborhood amenities, $\tilde{\alpha}_n^k$, the measures of residential tenure, $\beta_\tau^k \tilde{\tau}_x$, and the time-invariant components of the moving costs variables, MC^k, MC^{kc}, and $\boldsymbol{\beta}_d^{k\prime} \boldsymbol{d}(n_t, n_{t-1})$.[38] Our main concern now is that *changes* in the observed components of households' flow utilities are correlated with *changes* in unobserved neighborhood amenities and household expectational errors. We must therefore construct neighborhood-level instruments, z_n, for our endogenous regressors that are orthogonal to both of these components:

$$
\begin{aligned}
(13) \qquad 0 &= \mathbb{E}\left[z_n \, \Delta \nu_{xnn'\tilde{n}}^k\right] \\
&= \mathbb{E}\left[z_n \, \left(\Delta \xi_n^k + \Delta\, \tilde{e}(x, \bar{\omega}_t^{kc}, \bar{\omega}_{t'}^{kc})\right)\right] \\
&= \mathbb{E}\left[z_n \, \left(\Delta \xi_n^k + \Delta \left(e^{\bar{V}}(n, x, \bar{\omega}_t^{kc}, \bar{\omega}_{t'}^{kc}) - e^{\bar{V}}(n', x, \bar{\omega}_t^{kc}, \bar{\omega}_{t'}^{kc})\right)\right)\right],
\end{aligned}
$$

where, $e^{\bar{V}}(n, x, \bar{\omega}_t^{kc}, \bar{\omega}_{t'}^{kc})$ is the difference between the realized type-k ex-ante continuation value and type-k households' expectations of these continuation values, integrated over the potential realizations of the household-level states:

$$
e^{\bar{V}}(n, x, \bar{\omega}_t^{kc}, \bar{\omega}_{t'}^{kc}) \equiv \sum_{x'} \left(\bar{V}(x', \bar{\omega}_{t'}^{kc}) - \mathbb{E}_{\bar{\omega}'|\bar{\omega}_t^{kc}}\left[\bar{V}(x', \bar{\omega}')|\bar{\omega}_t^{kc}\right]\right) f^x\left(x'|n, x_n, \bar{\omega}_t^{kc}\right).
$$

In addition to being orthogonal to changes in unobserved neighborhood amenities, equation 13 shows our instruments must also be mean independent of changes in households' expectational errors. Recall assumption (c) which states that households have rational expectations over the evolution of the model's state variables. A corollary of this assumption is that the contents of households' information sets at time t are mean independent of their expectational errors (Kalouptsidi, Scott, and Souza-Rodrigues 2021). Conversely, elements of households' future information sets that cannot be predicted from their period-t information sets will be correlated with their expectational errors. For this reason, our instruments must not predict future values of $\bar{\omega}_t^{kc}$ in a way that cannot simultaneously be predicted using information in households' period-t information sets. To see why, consider an instrument that shocks the neighborhood-n elements of $\omega_{t'}^{kc}$ for any $t \leq 2017$.[39] Assume this shock is uncorrelated with changes in unobserved neighborhood amenities but cannot be predicted with information in households' contemporaneous information sets, $\mathcal{I}_{it}$. If this instrument is relevant it will be mechanically

[38] We estimate the remaining time-invariant parameters in a final stage, estimating equation 11 conditional on the estimates from our differenced regressions and our first-step multinomial choice model. We assume that the effect of residential tenure on the likelihood of different residential paths is uncorrelated with unobserved neighborhood amenities.

[39] If the instrument shocks elements of $\omega_{t'}^{kc}$ for $t' > 2017$, we must also consider how the instrument affects the difference in expectational errors over time. The current example is sufficient to show that z_n must be constructed using variation that can be predicted from households' 2010 information sets.

correlated with the realized values of households' time-t ex-ante continuation values, $\bar{V}(x', \bar{\omega}_{t'}^{kc})$, but uncorrelated with households' time-t expectations, $\mathbb{E}_{\bar{\omega}'|\bar{\omega}_t^{kc}}\left[\bar{V}(x', \bar{\omega}')\middle|\bar{\omega}_t^{kc}\right]$, violating the exclusion restriction embodied in 13. For this reason, the instruments we detail immediately below are designed to predict changes in the CBSA-level state variables through 2011-2018 using only variation that can be predicted from households' 2010 information sets.

6. Identification

We now present the instrumental variables we use to identify low-income renters' preferences for our three endogenous variables, β_w^k, β_r^k, and β_A^k. Our first set of instruments predict changes in neighborhood-level job market access for our target population of low-income renters, helping to identify their preference for job market access, β_w^k. Our second set of instruments are similar but predict changes in neighborhood-level job market access for college graduates. By predicting neighborhood demand among college graduates, these instruments help identify preferences for endogenous amenities, β_A^k. Our third set of instruments interact predicted changes in neighborhood-level job market access for college graduates with the intensity of neighborhoods' ex-ante urban development.[40] Neighborhoods with an ex-ante high share of urban development are more likely to be on the inelastic portion of the local housing supply curve, making demand shocks more likely to translate into higher rents. These interaction terms are useful for identifying households' distaste for paying rent, β_r^k. To increase first stage power, we include a fourth set of instruments. These instruments use the proximity to other neighborhoods' shares of college graduates to help predict neighborhood demand among college graduates. These instruments are therefore particularly helpful in identifying households' preferences for endogenous amenities, β_A^k.

6.1. Job Market Access IV

Neighborhoods differ in their access to employment opportunities. Neighborhoods located near establishments with a high demand for skilled labor will be attractive to college graduates due to shorter expected commute times among this group, all else equal. A similar argument follows for low-income households and neighborhoods near establishments employing low-skill workers. Our job market access instruments predict changes in the desirability of neighborhoods based on changes to their expected commute times to employment opportunities. We construct these instruments in two steps. First, we define industry- and neighborhood-specific measures of job

[40]Specifically, we interact changes in job market access with the share of land in the census tract that is covered by urban development in 2011. We obtain these measures from Baum-Snow and Han (2023) who in turn construct them using data from the National Land Cover Database.

market access in a baseline year, separately for college graduates and low-income households. Second, we interact these baseline measures with national industry employment trends to predict changes in job market access that are plausibly uncorrelated with underlying trends in neighborhoods' exogenous amenities. We discuss how we construct these instruments as well as threats to identification below.

Instrument Construction. Our neighborhood-level measures of job market access can be formally defined in terms of each neighborhood n's access to employment in industry d at time t:

$$JMA_{ndt} = \sum_{m \in \mathcal{N}^c \backslash n} e^{-\eta^c \tau_{nm}} l_{mdt}, \tag{14}$$

where l_{mdt} is the number of jobs in workplace tract m and 6-digit NAICS industry d in time t, $\tau_{n,m}$ is the travel-time between tracts n and m, and η^c is a spatial decay parameter governing the importance of faraway jobs relative to closer jobs in determining a tract's employment access.

We obtain measures of $\tau_{n,m}$ for college graduates using reported commute destinations and times from our ACS data. However, since the number of neighborhood-pairs in each CBSA is large relative to the number of college graduates surveyed in the ACS, we follow Baum-Snow, Hartley, and Lee (2019) and estimate a simple forecasting model to predict $\tau_{n,m}$ for all neighborhood pairs in each CBSA. We obtain CBSA-specific measures of η^c by using our employer-employee linked data to estimate gravity equations derived from a workplace choice model á la Tsivanidis (2022). We detail both our forecasting model and workplace choice model with its resulting gravity equations in appendix B.3.[41]

We use these industry-specific measures of employment access to construct our job market access IVs:

$$\Delta \widetilde{JMA}_{n,t_0,t} = \sum_{d \in \mathcal{T}} \underbrace{\frac{JMA_{n,d,t_0} \theta^c_d}{\sum_{d'} JMA_{n,d',t_0} \theta^c_{d'}}}_{Share} \underbrace{\frac{L^c_{d,t} - L^c_{d,t_0}}{L^c_{d,t_0}}}_{Shift}, \tag{15}$$

where $L^c_{d,t}$ is national employment in industry d less employment in industry d located in neighborhood n's CBSA. $\mathcal{T}$ denotes the set of tradable of industries we use to predict changes in job market access.[42] As we do not observe the educational level or race of workers in our

[41] Travel times and spatial decay parameters are defined separately for college graduates and low-income renters.

[42] We define our set of tradable industries using trade costs for 6-digit NAICS manufacturing and service industries, as estimated in Gervais and Jensen (2019). We label an industry as tradable if its estimated trade costs are in the bottom three quartiles of the manufacturing and service industries analyzed by Gervais and Jensen (2019). We find this threshold ensures sufficient first-stage power while excluding industries whose establishment locations

establishment-level data, we scale our industry-level measures of job market access by the share of workers employed in each industry and in each state that have a college degree, θ_d^c. Remember that analogous IVs are constructed for low-income renter households.

For all our job market access instruments, we select $t_0 = 2002$ and $t = 2007$. As our discussion around households' expectational errors in section 5.3 highlights, our instruments may not use information outside of households' information sets to shift the endogenous variables. If the instruments predict changes in the endogenous variables that households do not - on average - expect, then the instruments will be correlated with their expectational errors. By setting $t_0 = 2002$ and $t = 2007$, we rely on serial correlation in the endogenous variables to ensure our instruments remain relevant. Our choice of $t_0 = 2002$ is motivated by the fact that the US Census Bureau's Economic Census occurs on the years ending in 2 and 7 (Chow et al. 2021). The allocation of firm employment data across establishments is most accurate in these years, increasing the precision of our baseline shares.[43] We find that first-stage power for our job market access instruments are then maximized when we choose $t = 2007$; local labor demand shocks induced by the Great Recession don't appear to influence households' location choices throughout 2011-18.

Identifying Assumptions. Recent studies on shift-share instruments show how the exclusion restriction can hold with either conditionally exogenous shares (Goldsmith-Pinkham, Sorkin, and Swift 2020) or with conditionally exogenous shifts (Borusyak, Hull, and Jaravel 2022). The shares in equation 15 correspond to neighborhoods' baseline commute-time-discounted exposure to employment in each industry. The shifts correspond to the national employment growth rate in each industry. As it is less plausible that establishments' baseline neighborhood locations are unrelated to underlying trends in nearby neighborhoods' exogenous amenities, we argue that identification comes from the conditional exogeneity of national industry employment shifts.

Borusyak, Hull, and Jaravel (2022) show that three conditions are together sufficient to ensure our employment "shifts" are conditionally exogenous. First, establishments in industries with nationwide employment shocks (positive or negative) must not be concentrated near neighborhoods experiencing trends (positive or negative) in their unobserved exogenous amenities. Second, no small subset of industries may comprise a large portion of the baseline shares. Third, industries' national employment shifts must be mutually uncorrelated given trends in

are likely endogenous to the spatial sorting within cities, such as local retailers. We discuss our choice to focus on tradable industries below.

[43]We considered using 1997 as our baseline year, but this would have required manually geocoding establishments' addresses since the LBD had not started reporting establishments' census tracts at that time.

unobserved amenities and baseline shares.[44]

We consider a number of threats to identification. First, researchers have argued that throughout our analysis period there was a general trend toward suburbanization among low-income households irrespective of gentrification (Bartik and Mast 2023). If establishments concentrated near suburban neighborhoods that primarily employed low-skilled workers were overrepresented in industries experiencing negative nationwide employment shocks, then we may mistake a secular migration trend as a distaste for market access. To account for this possibility, we residualize $\Delta \widetilde{JMA}_{n,t_0,t}$ on measures of proximity to the metro division's CBD. These measures include a quadratic in the physical distance between the CBD and neighborhood n's centroid, a quadratic in the population-weighted distance between the CBD and neighborhood n's centroid, and fixed effects for five equally-sized concentric rings that are centered at the CBD. The concentric rings are measured in population-weighted distance. Together, these measures ensure that our job market access instruments induce variation in the endogenous variables among neighborhoods that are equidistant from each metro division's CBD.

A second threat to identification is that changing consumer preferences may jointly influence households' location choices *and* industry employment trends. For example, changing preferences for different types of non-tradable services may simultaneously influence employment in those industries and households' within-CBSA residential location decisions. To avoid this concern, we construct $\Delta \widetilde{JMA}_{n,t_0,t}$ using only employment shifts in *tradable* 6-digit NAICS industries.[45] This ensures employment trends are not caused by households' residential location choices. We also construct our national employment shifters excluding employment in the CBSA for which we are predicting changes in job market access.

A third threat to identification relates to the secular decline in manufacturing (NAICS 31-33) employment throughout 2002-07 (Autor, Dorn, and Hanson 2013). Such correlated employment shocks conditional on baseline shares and unobserved amenities threaten the consistency of our estimates. To account for correlated employment shocks within the manufacturing sector, we residualize $\Delta \widetilde{JMA}_{n,t_0,t}$ on neighborhoods' baseline exposure to manufacturing employment. Identification then requires industry employment shifts *within* the manufacturing sector are mutually uncorrelated conditional on baseline shares and unobserved amenities, a significantly

[44]The first condition can be represented formally as $\mathbb{E}[g_d | \{\bar{\Delta\nu}_d\}_d, \{s_d\}_d] = \mu$, the second condition as $\mathbb{E}\left[\sum_d s_d^2\right] \to 0$, and the third condition as $\text{Cov}\left(g_d, g_{d'} | \{\bar{\Delta\nu}_d\}_d, \{s_d\}_d\right) = 0 \forall (d, d')$, where $s_d = \sum_n s_{n,d} = \sum_n \frac{JMA_{n,d,t_0}\theta_d^c}{\sum_{d'} JMA_{n,d',t_0}\theta_{d'}^c}$, $g_d = \frac{L_{d,t}^c - L_{d,t_0}^c}{L_{d,t_0}^c}$, and $\bar{\Delta\nu}_d = \frac{\sum_{nn'\tilde{n}} s_{n,d}\Delta\nu_{nn'\tilde{n}}}{s_d}$.

[45]By restricting to tradable industries, the effective "shares" in $\Delta \widetilde{JMA}_{n,t_0,t}$ do not sum to 1 across industries and within neighborhoods. To control for the possibility that neighborhoods near high concentrations of tradable industry establishments are systematically exposed to increasing/decreasing exogenous neighborhood amenities, we additionally control for neighborhoods' exposure to the total share of tradable employment (Borusyak, Hull, and Jaravel 2022).

less stringent requirement (Borusyak, Hull, and Jaravel 2022).

6.2. Proximity IV

Guerrieri, Hartley, and Hurst (2013) show that during a positive city-wide employment demand shock, among low-income neighborhoods it is those closest to other high-income neighborhoods that experience the greatest home price appreciation - a proxy for gentrification. Motivated by these findings, we construct a neighborhood-level and distance-weighted measure of proximity to other neighborhoods' share of college graduates. We then interact this measure with CBSA-wide Bartik labor demand shocks constructed using initial CBSA-wide shares of college graduates in tradable 6-digit NAICS industries. Again, though our instrument construction is novel, we are not the first to use proximity to other high-income neighborhoods as an instrument for gentrification (Brummet and Reed (2021); Glaeser, Luca, and Moszkowski (2023)).

Instrument Construction. We first form neighborhood-level distance-weighted measures of proximity to other neighborhoods' shares of college graduates in the same CBSA. We then interact these neighborhood-level measures with shocks to CBSA-wide tradable industry employment:

$$\Delta\,\widetilde{\text{Prox}}_{n,t_0,t',t} = \sum_{d\in\mathcal{T}} \underbrace{\sum_{m\in\mathcal{N}^c\setminus n} e^{\rho\tau_{n,m}}\frac{Coll_{m,t_0}}{Pop_{m,t_0}}}_{Proximity} \cdot \underbrace{\frac{l_{d,t'}\cdot\theta^c_d}{\sum_{d'} l_{d',t'}\cdot\theta^c_{d'}}}_{Share}\underbrace{\frac{L^c_{d,t}-L^c_{d,t'}}{L^c_{d,t'}}}_{Shift},$$

where $\frac{Coll_{m,t_0}}{Pop_{m,t_0}}$ is the share of residents in neighborhood m who are college graduates, and where $l_{d,t'} = \sum_n l_{d,n,t'}$. The variables $l_{d,n,t'}$, $L^c_{d,t}$, and the parameters $\theta^c_{d'}$ are defined as before. ρ is a spatial decay parameter governing the importance of faraway neighborhood college graduate shares relative to closer neighborhood shares. As $\rho\to\infty$, only the neighborhoods immediately adjacent to neighborhood n matter for determining the value of $\Delta\,\widetilde{\text{Prox}}_{n,t_0,t}$. Conversely, as $\rho\to 0$ the instrument loses all relevance as every neighborhood matters equally in determining $\Delta\,\widetilde{\text{Prox}}_{n,t_0,t}$, ensuring $\Delta\,\widetilde{\text{Prox}}_{n,t_0,t}$ is constant within each CBSA. Since we don't have a good prior for ρ, we calibrate its value using k-fold cross-validation in a set of first-stage regressions, regressing $\text{Gent}_{n,t_0,t}$ on $\Delta\,\widetilde{\text{Prox}}_{n,t_0,t}$ and selecting the value of ρ that best predicts changes in the neighborhood share of college graduates.

Unlike local labor demand, industry wide employment shocks throughout the Great Recession are very predictive of future *CBSA-wide* labor demand. For that reason, we use employment shifts between $t' = 2007$ and $t = 2010$ to construct our Bartik shift-share, but construct the proximity shares in $t_0 = 2002$ for reasons we discuss below.

Identification. Identification here proceeds differently to our job market access instruments. Because our final estimating equations include CBSA-wide fixed effects, identifying variation must come from within CBSAs and so rules out identification from the CBSA-wide shift-shares alone. Identification instead proceeds from the interaction between the "proximity" term and the Bartik shift-share.

Note that neighborhoods' proximity to other neighborhoods with a high share of college graduates is likely correlated with underlying trends in exogenous amenities. To account for this possibility, we take two steps. First, we use lagged shares of neighborhoods' proximity to other neighborhoods with a high share of college graduates. Second, we residualize $\Delta\,\widetilde{\text{Prox}}_{n,t_0,t',t}$ on the proximity terms themselves, so that identification comes solely from the *interaction* between the shift-shares and the proximity terms. Identification thus proceeds analogously to a difference-in-difference estimator: we compare differences in gentrification between neighborhoods near and far already gentrified tracts in CBSAs experiencing large labor demand shocks to differences in gentrification between neighborhoods near and far already gentrified tracts in CBSAs not experiencing large labor demand shocks.

6.3. Moment Conditions

The final collection of instruments are:

$$z_{1,n} = \Delta\,\widetilde{JMA}^{Coll}_{n,02,07} \tag{16}$$

$$z_{2,n} = \Delta\,\widetilde{JMA}^{Low-Income}_{n,02,07} \tag{17}$$

$$z_{3,n} = \Delta\,\widetilde{JMA}^{Coll}_{n,02,07} \cdot \varphi_n, \tag{18}$$

$$z_{4,n} = \Delta\,\widetilde{\text{Prox}}_{n,02,07,10}, \tag{19}$$

where φ_n are neighborhood-level measures of the share of land in the census tract covered by urban development in 2011. The super-scripts $Coll$ and $Low-Income$ refer to the group for whom the instrument is constructed.[46] We estimate the moment restriction in equation 13 via linear GMM separately for each type-k household:

$$\mathbb{E}\left[z_n\,\Delta\nu^k_{xnn'\tilde{n}}\right] = 0.$$

[46]We estimate the commute elasticities and the employment share parameters for Black and Non-Black low-income households jointly when constructing $\Delta\,\widetilde{JMA}^{Low-Income}_{n,02,07}$. This ensures the same variation identifies both type-k household's preferences.

A corresponding observation in this moment restriction is a set of residential paths for a type-k household with individual state, x. Each type-k household of city c has $2 \cdot (N^c + 1)$ initial states in period $t - 1$, N^c possible neighborhood choices in that same period, and $N^c - 1$ possible neighborhood choices in period t, implying $\sum_c 2 \cdot (N^c + 1) \cdot N^c \cdot (N^c - 1)$ observations for for both Black and Non-Black low-income renter households. Since the urban cores of our largest CBSAs contain roughly two thousand census tracts, the number of potential residential paths reach into the tens of billions. To ease the computational burden, we restrict the set of residential paths we analyze. Specifically, we restrict the set of residential paths to those that start from neighborhoods with the highest shares of type-k low-income renters in their CBSA. The cutoff for "highest shares" varies across CBSAs. For CBSAs with over 500 Census tracts in their urban core, we select the top ten percent of tracts in terms of their share of type-k low-income renters. For CBSAs with under 500 Census tracts in their urban core, we select the 50 highest tracts in terms of their share of type-k low-income renters. This choice ensures a minimum number of census tracts from each CBSA inform our estimates. We report our estimates in Table 4.

6.4. Parameter Estimates

The leftmost columns of Panel A in Table 4 report parameter estimates for households' valuations of neighborhood characteristics while the same columns in Panel B report estimates on households' moving frictions. Since logit models only identify relative changes in welfare, the rightmost columns in Table 4 translate the estimates into households' willingness to pay measured in annual rents. The annual rent values for college share (β_A) and market access (β_w) reflect the extra annual rent payments households would incur to reside in a neighborhood with a 10% higher share of college graduates or market access, respectively.[47]

The interpretation of the annual rent valuations differ across the moving friction parameter estimates. The annual rent values for physical move distance (β_d and β_d^{Sq}) represent the amount of annual rents households would be willing to pay to move to a census tract one mile closer to their origin neighborhood, conditional on moving. The annual rent values for social distance (β_s and β_s^{Sq}) represent the amount of annual rents households would be willing to pay to move to a census tract 10 percentage points more similar to their current census tract in terms of neighborhood college shares, again conditional on moving.[48] The fixed moving cost annual rent valuations represent how much in annual rents households would be willing to pay to avoid moving in any one year. The high residential tenure annual rent valuations represent how much

[47] We use the average neighborhood rent levels for Black and Non-Black households in Table A1 as the basis of the percentage change in rents.

[48] The physical distance value of rent measures were calculated based on an initial move of 0 miles. The social distance value of rent measures were calculated assuming the average baseline share of college graduates for both Black and Non-Black households, as reported in Table **??**.

the fixed cost of moving increases each additional year the household remains in their current neighborhood.

Our parameter estimates all have the expected sign, though Black households have a surprisingly low valuation for market access.[49] One explanation for Black households' low valuation for market access is that the variation in market access induced by our job market access instruments is concentrated in industries primarily employing Non-Black workers; such variation would not induce migration responses from households not employed in these industries, suggesting no preference for market access. Future iterations of the paper will test the robustness of our parameter estimates to skill and race-specific measures of job market access.

Our parameter estimates reveal how costly it is to move across neighborhoods, even for renter households, and especially for those with a high level of residential tenure. Low-income Black renter households' within-CBSA migration decisions suggest that the fixed cost of moving neighborhoods is over \$3,000, and that this cost increases by around \$600 with every additional year of residential tenure. Surprisingly, the social distance between neighborhoods plays little role in households' residential migration decisions. One explanation for this result is that low-income households initially residing in gentrifying tracts move to tracts with a lower share of college graduates, conditional on moving. Our model would interpret these moves as revealing a preference for social distance. A different model specification may incorporate asymmetric moving costs, capturing the distinct cost of moving into a neighborhood with a higher share of college graduates than one's origin tract.

7. Welfare Analysis

With households' demand parameter estimates in hand, we may now examine the effect of gentrification on the welfare of low-income incumbent residents through 2010-2019. To this end, we take our estimated parameters and compute the ...

References

Abadie, Alberto, Susan Athey, Guido Imbens, and Jeffrey Wooldridge. 2023. "When should you adjust your standard errors for clustering?" *The Quarterly Journal of Economics* 138 (1): 1–35.

Abowd, John, Bryce Stephens, Lars Vilhuber, Fredrik Andersson, Kevin McKinney, Marc Roemer, and Simon Woodcock. 2009. "The LEHD Infrastructure Files and the Creation of the Quarterly Workforce

[49] While Black households appear to value some social distance conditional on moving, they have a distaste for moves that involve significant differences in neighborhood educational composition.

TABLE 4. Parameter Estimates

	Estimates		Value in Annual Rent	
	Black	Non-Black	Black	Non-Black
Panel A: Neighborhood Characteristics				
Rents (β_r)	-10.31***	-26.40***	–	–
	(1.494)	(3.341)		
College Share (β_A)	12.61***	7.343***	$1,224	$312
	(0.390)	(1.437)		
Market Access (β_w)	0.122	20.04***	$12	$851.4
	(1.635)	(3.487)		
Panel B: Moving Frictions				
Distance (β_d)	-0.2782***	-.3091***	-$270	-$130
	(0.0004)	(0.0003)		
Distance Squared (β_d^{Sq})	0.0039***	0.0044***		
	(0.0000)	(0.0000)		
Social Distance (β_s)	0.1335***	-0.1751***	$5	-$15
	(0.0108)	(0.0084)		
Social Distance Squared (β_s^{Sq})	-0.5598***	-0.4096***		
	(0.0074)	(0.0060)		
Within-CBSA Fixed Moving Cost (MC)	-4.558***	-4.318***	-$3,578	-$1,692
	(0.000)	(0.0000)		
Outside-CBSA Fixed Moving Cost (MC^c)	-4.458***	-4.621***	-$3,515	-$1,801
	(0.000)	(0.0000)		
High Residential Tenure ($\bar{\tau}$)	0.612***	0.551***	$612	$237
	(0.0002)	(0.0001)		
F-Stat	63.35	27.92		
N (1,000s)	2,573,000	10,630,000		

Notes: An observation is a set of residential paths described in section B.1. The number of residential paths in our analysis differs across Black and Non-Black low-income renter households because we predict fewer conditional choice probabilities for Black households in the first-step of the analysis, as discussed in footnote 36.

Indicators." In *Producer Dynamics: New Evidence from Micro Data,* edited by Timothy Dunne, Bradford Jensen, and Mark Roberts, chap. 5, 149–230: University of Chicago Press.

Abramson, Boaz. 2023. "The Equilibrium Effects of Eviction and Homelessness Policies." *Working Paper*.

Aguirregabiria, Victor, and Arvind Magesan. 2013. "Euler Equations for the Dynamic Estimation of Discrete Choice Structural Models." *Structural Econometric Methods: Advances in Econometrics* 31: 3–44.

Almagro, Milena, and Tomás Domínguez-Lino. 2022. "Location Sorting and Endogenous Amenities: Evidence from Amsterdam." *Working Paper*.

Angrist, Joshua D, and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion.*: Princeton university press.

Arcidiacono, Peter, and Robert Miller. 2011. "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity." *Econometrica* 79 (6): 1823–1867.

Autor, David H, David Dorn, and Gordon H Hanson. 2013. "The China syndrome: Local labor market effects of import competition in the United States." *American economic review* 103 (6): 2121–2168.

Autor, David H, Christopher J Palmer, and Parag A Pathak. 2017. "Gentrification and the amenity value of crime reductions: Evidence from rent deregulation."Technical report.

Balboni, Clare, Gharad Bryan, Melanie Morten, and Bilal Siddiqi. 2020. "Transportation, gentrification, and urban mobility: The inequality effects of place-based policies." *Working Paper*.

Bartik, Alexander, and Evan Mast. 2023. "Black Suburbanization: Causes and Consequences of a Transformation of American Cities." *Working Paper*.

Baum-Snow, Nathaniel, and Lu Han. 2023. "The Microgeography of Housing Supply." *Working Paper*.

Baum-Snow, Nathaniel, and Daniel Hartley. 2020. "Accounting for central neighborhood change, 1980-2010." *Journal of Urban Economics* 117: 103228.

Baum-Snow, Nathaniel, Daniel Hartley, and Kwan Ok Lee. 2019. "The Long-Run Effects of Neighborhood Change on Incumbent Families." *Working Paper*.

Bayer, Patrick, Fernando Ferreira, and 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, and Christopher Timmins. 2016. "A Dynamic Model of Demand for Houses and Neighborhoods." *Econometrica* 84 (3): 893–942.

Bergman, Peter, Raj Chetty, Stefanie DeLuca, Nathaniel Hendren, Lawrence Katz, and Christopher Palmer. 2023. "Creating Moves to Opportunity: Experimental Evidence on Barriers to Neighborhood Choice." *Working Paper*.

Berkes, Enrico, and Ruben Gaetani. 2023. "Income Segregation and the Rise of the Knowledge Economy." *American Economic Journal: Applied Economics* 15 (2): 69–102.

Berry, Christopher. 2021. "Reassessing the Property Tax." *Working Paper*.

Bishop, Kelly C. 2012. "A Dynamic Model of Location Choice and Hedonic Valuation." *Mimeo, Washington University in St. Louis*: 1–38.

Bishop, Kelly C, and Alvin D Murphy. 2019. "Valuing Time-Varying Attributes Using the Hedonic Model: When is a Dynamic Approach Necessary?" *Review of Economics and Statistics* 101 (1): 134–145.

Böhlmark, Anders, and Alexander Willén. 2020. "Tipping and the Effects of Segregation." *American Economic Journal: Applied Economics* 12 (1): 318–47.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2022. "Quasi-Experimental Shift-Share Research Designs." *Review of Economic Studies* 89 (1): 181–213.

Boustan, Leah Platt. 2010. "Was Postwar Suburbanization "White Flight"? Evidence from the Black Migration." *The Quarterly Journal of Economics* 125 (1): 417–443.

Brock, William A, and Steven N Durlauf. 2002. "A Multinomial-Choice Model of Neighborhood Effects."

American Economic Review, P&P 92 (2): 298–303.
Brueckner, Jan. 1987. “The structure of urban equilibria: A unified treatment of the Muth-Mills model.” *Handbook of Regional and Urban Economics* 2: 821–845.
Brueckner, Jan K, Jacques-François Thisse, and Yves Zenou. 1999. “Why is central Paris rich and downtown Detroit poor?: An amenity-based theory.” *European Economic Review* 43 (1): 91–107.
Brummet, Quentin, and Davin Reed. 2021. “The Effects of Gentrification on Incumbent Residents.” *Working Paper*.
Card, David, Alexandre Mas, and Jesse Rothstein. 2008. “Tipping and the Dynamics of Segregation.” *Quarterly Journal of Economics* 123 (1): 177–218.
Chow, Melissa, Teresa Fort, Christopher Goetz, Nathan Goldschlag, James Lawrence, Elisabeth Perlman, Marther Stinson, and Kirk White. 2021. “Redesigning the Longitudinal Business Database.” *NBER Working Paper* (28839).
Chyn, Eric, and Lawrence F Katz. 2021. “Neighborhoods Matter: Assessing the Evidence for Place Effects.” *Journal of Economic Perspectives* 35 (4): 197–222.
Collinson, Robert, John Eric Humphries, Nicholas Mader, Davin Reed, Daniel Tannenbaum, and Winnie van Dijk. 2023. “Eviction and Poverty in American Cities.” *Quarterly Journal of Economics*.
Correia, Sergio, Paulo Guimarães, and Tom Zylkin. 2020. “Fast Poisson estimation with high-dimensional fixed effects.” *The Stata Journal* 20 (1): 95–115.
Couture, Victor, Cécile Gaubert, Jessie Handbury, and Erik Hurst. 2023. “Income Growth and the Distributional Effects of Urban Spatial Sorting.” *The Review of Economic Studies*.
Couture, Victor, and Jessie Handbury. 2020. “Urban Revival in America.” *Journal of Urban Economics* 119: 103267.
Couture, Victor, and Jessie Handbury. 2023. “Neighborhood Change, Gentrification, and the Urbanization of College Graduates.” *Journal of Economic Perspectives* 37 (2): 29–52.
Cullen, Julie Berry, and Steven D Levitt. 1999. “Crime, Urban Flight, and the Consequences for Cities.” *The Review of Economics and Statistics* 81 (2): 159–169.
Curci, Federico, and Federico Masera. 2023. “Flight from Urban Blight: Lead Poisoning, Crime, and Suburbanization.” *The Review of Economics and Statistics*: 1–45.
Curci, Federico, and Hasin Yousef. 2022. “Gentrifying Cities, Amenities and Income Segregation: Evidence from San Francisco.” *Working Paper*.
Davis, Morris, Jesse Gregory, and Daniel Hartley. 2023. “Preferences over the Racial Composition of Neighborhoods: Estimates and Implications.” *Working Paper*.
Davis, Morris, Jesse Gregory, Daniel Hartley, and Kegon Tan. 2021. “Neighborhood Effects and Housing Vouchers.” *Quantitative Economics* 12 (4): 1307–1346.
Dekle, Robert, Jonathan Eaton, and Samuel Kortum. 2008. “Global Rebalancing with Gravity: Measuring the Burden of Adjustment.” *IMF Staff Papers* 55 (3): 511–540.
Derenoncourt, Ellora. 2022. “Can You Move to Opportunity? Evidence from the Great Migration.” *American Economic Review* 112 (2): 369–408.
Diamond, Rebecca. 2016. “The Determinants and Welfare Implications of US Workers’ Diverging Location Choices by Skill: 1980-2000.” *American Economic Review* 106 (3).
Diamond, Rebecca, and Cecile Gaubert. 2022. “Spatial Sorting and Inequality.” *Annual Review of Economics* 14: 795–819.
Diamond, Rebecca, Timoth McQuade, and Franklin Qian. 2018. “The Effects of Rent Control Expansion on Tenants, Landlords, and Inequality: Evidence from San Francisco.” *NBER Working Paper* (24181).
Ding, Lei, and Jackelyn Hwang. 2020. “Effects of Gentrification on Homeowners: Evidence from a Natural

Experiment." *Regional Science and Urban Economics* 83: 103536.
Ding, Lei, Jackelyn Hwang, and Eileen Divringi. 2016. "Gentrification and Residential Mobility in Philadelphia." *Regional Science and Urban Economics* 61: 38–51.
Donaldson, Dave, and Richard Hornbeck. 2016. "Railroads and American Economic Growth: A "Market Access" Approach." *The Quarterly Journal of Economics* 131 (2): 799–858.
Dragan, Kacie, Ingrid Gould Ellen, and Sherry Glied. 2020. "Does Gentrification Displace Poor Children and their Families? New evidence from Medicaid Data in New York City." *Regional Science and Urban Economics* 83: 103481.
Edlund, Lena, Cecilia Machado, and Maria Sviatschi. 2022. "Gentrification and the Rising Returns to Skill." *Economica* 89 (354): 258–292.
Ellen, Ingrid Gould, Keren Mertens Horn, and Davin Reed. 2019. "Has Falling Crime Invited Gentrification?" *Journal of Housing Economics* 46: 101636.
Ellen, Ingrid Gould, and Katherine O'Regan. 2010. "Crime and urban flight revisited: The effect of the 1990s drop in crime on cities." *Journal of Urban Economics* 68 (3): 247–259.
Ellen, Ingrid Gould, and Katherine M O'Regan. 2011. "How Low Income Neighborhoods Change: Entry, Exit, and Enhancement." *Regional Science and Urban Economics* 41 (2): 89–97.
Fee, Klye, and Daniel Hartley. 2013. "The Relationship Between City Center Density and Urban Growth or Decline." In *Revitalizing American Cities*, edited by Kimberly Zeuli and Susan Wachter, chap. 3, 45–64: University of Pennsylvania Press.
Freeman, Lance. 2005. "Displacement or Succession? Residential Mobility in Gentrifying Nneighborhoods." *Urban Affairs Review* 40 (4): 463–491.
Freeman, Lance, and Frank Braconi. 2004. "Gentrification and Displacement New York City in the 1990s." *Journal of the American planning association* 70 (1): 39–52.
Fu, Ellen. 2022. "The Financial Burdens of Property Taxes: Evidence from Philadelphia." *Working Paper*.
Gabriel, Stuart, and Gary Painter. 2020. "Why Affordability Matters." *Regional Science and Urban Economics* 80: 103378.
Gans, Herbert J. 1982. *Urban Villagers*.: Simon and Schuster.
Gendron-Carrier, Nicolas. 2023. "Prior Work Experience and Entrepreneurship: The Careers of Young Entrepreneurs." *IZA Discussion Paper* (16145).
Gervais, Antoine, and J. Bradford Jensen. 2019. "The tradability of services: Geographic concentration and trade costs." *Journal of International Economics* 118: 331–350.
Glaeser, Edward, Matthew Kahn, and Jordan Rappaport. 2008. "Why do the poor live in cities? The role of public transportation." *Journal of Urban Economics* 63: 1–24.
Glaeser, Edward, Jed Kolko, and Albert Saiz. 2001. "Consumer City." *Journal of Economic Geography* 1: 27–50.
Glaeser, Edward L. 2020. "Urbanization and its Discontents." *Eastern Economic Journal* 46: 191–218.
Glaeser, Edward L, and Matthew E Kahn. 2001. "Decentralized Employment and the Transformation of the American City.": 1–63.
Glaeser, Edward L, and Erzo F P Luttmer. 2003. "The Misallocation of Housing Under Rent Control." *American Economic Review* 93 (4): 1027–1046.
Glaeser, Edward, Michael Luca, and Eric Moszkowski. 2023. "Gentrification and retail churn: Theory and evidence." *Regional Science and Urban Economics* 100: 103879.
Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift. 2020. "Bartik Instruments: What, When, Why, and How." *American Economic Review* 110 (8): 2586–2624.
Groote, Olivier De, and Frank Verboven. 2019. "Subsidies and Time Discounting in New Technology

Adoption: Evidence from Solar Photovoltaic Systems." *American Economic Review* 109 (6): 2137–72.
Guerrieri, Veronica, Daniel Hartley, and Erik Hurst. 2013. "Endogenous Gentrification and Housing Price Dynamics." *Journal of Public Economics* 100: 45–60.
Guimarães, Paulo, Octávio Figueirdo, and Douglas Woodward. 2003. "A Tractable Approach to the Firm Location Decision Problem." *Review of Economics and Statistics* 85 (1): 201–204.
Hoelzlein, Mattias. 2023. "Two-Sided Sorting and Spatial Inequality in Cities." *Working Paper*.
Hotz, Joseph, and Robert Miller. 1993. "Conditional Choice Probabilities and the Estimation of Dynamic Models." *The Review of Economic Studies* 60 (3): 497–529.
Hwang, Jackelyn, and Jeffrey Lin. 2016. "What Have We Learned About the Causes of Recent Gentrification?" *Cityscape* 18 (3): 9–26.
Jackson, Kenneth T. 1987. *Crabgrass Frontier: The Suburbanization of the United States.*: Oxford University Press.
Jargowsky, Paul A. 2009. "Immigrants and Neighbourhoods of Concentrated Poverty: Assimilation or Stagnation?" *Journal of Ethnic and Migration Studies* 35 (7): 1129–1151.
Kain, John. 1962. "The journey-to-work as a determinant of residential location." *Papers of the Regional Science Association* 9 (1): 137–160.
Kain, John F. 1968. "Housing Segregation, Negro Employment, and Metropolitan Decentralization." *The Quarterly Journal of Economics* 82 (2): 175–197.
Kalouptsidi, Myrto, Paul Scott, and Eduardo Souza-Rodrigues. 2021. "Linear IV regression estimators for structural dynamic discrete choice models." *Journal of Econometrics* 222: 778–804.
Kennan, John, and James R Walker. 2011. "The Effect of Expected Income on Individual Migration Decisions." *Econometrica* 79 (1): 211–251.
Kleinbaum, David G, and Mitchel Klein. 1996. *Survival Analysis: A Self-Learning Text.*: Springer.
Kuminoff, Nicolai V, V Kerry Smith, and Christopher Timmins. 2013. "The New Economics of Equilibrium Sorting and Policy Evaluation Using Housing Markets." *Journal of Economic Literature* 51 (4): 1007–1062.
Le Barbanchon, Thomas, Roland Rathelot, and Alexandra Roulet. 2021. "Gender Differences in Job Search: Trading off Commute against Wage." *The Quarterly Journal of Economics* 136 (1): 381–426.
LeRoy, Stephen, and Jon Sonstelie. 1983. "Paradise lost and regained: Transportation innovation, income, and residential location." *Journal of Urban Economics* 13: 67–89.
Lester, T William, and Daniel A Hartley. 2014. "The Long Term Employment Impacts of Gentrification in the 1990s." *Regional Science and Urban Economics* 45: 80–89.
Manson, Steven, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles. 2022. "IPUMS National Historical Geographic Information System: Version 17.0 [dataset]. Minneapolis, MN: IPUMS.."
Meltzer, Rachel, and Pooya Ghorbani. 2017. "Does Gentrification Increase Employment Opportunities in Low-Income Neighborhoods?" *Regional Science and Urban Economics* 66: 52–73.
Mieszkowski, Peter, and Edwin S Mills. 1993. "The Causes of Metropolitan Suburbanization." *Journal of Economic Perspectives* 7 (3): 135–147.
Miller, Conrad. 2021. "When Work Moves: Job Suburbanization and Black Employment." *Working Paper*.
Moreno-Maldonado, Ana, and Clara Santamaria. 2022. "Delayed Childbearing and Urban Revival." *Working Paper*.
Moretti, Enrico. 2013. "Real Wage Inequality." *American Economic Journal: Applied Economics* 5 (1): 65–103.
Notowidigdo, Matthew J. 2020. "The Incidence of Local Labor Demand Shocks." *Journal of Labor Economics* 38 (3): 687–725.
Redding, Stephen J, and Esteban Rossi-Hansberg. 2017. "Quantitative spatial economics." *Annual Review of Economics* 9: 21–58.

Roback, Jennifer. 1982. "Wages, Rents, and the Quality of Life." *Journal of Political Economy* 90 (6): 1257–1278.

Rosen, Sherwin. 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy* 82 (1): 34–55.

Rust, John. 1987. "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher." *Econometrica* 55 (5): 999–1033.

Scott, Paul. 2013. "Dynamic Discrete Choice Estimation of Agricultural Land Use." *Discussion Paper, New York University*.

Su, Yichen. 2022. "The Rising Value of Time and the Origin of Urban Gentrification." *American Economic Journal: Economic Policy* 14 (1): 402–439.

Traiberman, Sharon. 2019. "Occupations and Import Competition: Evidence from Denmark." *American Economic Review* 109 (12): 4260–4301.

Tsivanidis, Nick. 2022. "Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio." *Working Paper*.

Vigdor, Jacob L. 2002. "Does Gentrification Harm the Poor?" *Brookings-Wharton Papers on Urban Affairs*: 133–182.

Vigdor, Jacob L. 2010. "Is urban decay bad? Is urban revitalization bad too?" *Journal of Urban Economics* 68 (3): 277–289.

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

Wilson, William Julius. 1987. *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy.*: University of Chicago Press.

Appendix A. Aggregate Neighborhood Change Since 2000

Appendix B. Estimation Appendix

B.1. Deriving our Estimating Equation

Consider the set of residential choices detailed in section 5.2. Given these residential choices, we start the derivation of our moment restrictions with an application of the Hotz-Miller inversion, which amounts to differencing equation 6 across the two neighborhood choices in period t, n and n':

$$\begin{aligned}\ln\left(\frac{p^k_{xnt}}{p^k_{xn't}}\right) &= v^k_{xnt} - v^k_{xn't}\\ &= \bar{u}^k_{xnt} - \bar{u}^k_{xn't} + \delta\left(\mathbb{E}\left[\bar{V}^k_{xnt}\right] - \mathbb{E}\left[\bar{V}^k_{xn't}\right]\right),\end{aligned} \tag{A1}$$

where the expectation operator is with respect to both the observable household-level and all the city-specific state variables. By assumption (a), we can write these expectations as

$$\begin{aligned}\mathbb{E}\left[\bar{V}^k_{xnt}\right] &= \sum_{x'}\int_{\bar{\omega}'}\bar{V}^k_{xnt}dF^{\bar{\omega}}\left(\bar{\omega}'|\bar{\omega}\right) f^x\left(x'|n, x_{nt}, \bar{\omega}^{kc}_t\right)\\ &= \sum_{x'}\mathbb{E}_{\bar{\omega}'|\bar{\omega}^{kc}_t}\left[\bar{V}(x', \bar{\omega}')\middle|\bar{\omega}^{kc}_t\right] f^x\left(x'|n, x_{nt}, \bar{\omega}^{kc}_t\right)\end{aligned}$$

where x' and $\bar{\omega}'$ denote the next period values for x and $\bar{\omega}$. We can also replace the expectation of the ex-ante continuation values with respect to the city-specific state variables with their realized counterparts and an expectational error defined in equation 8:

$$\mathbb{E}\left[\bar{V}^k_{xnt}\right] = \sum_{x'}\bar{V}(x', \bar{\omega}^{kc}_{t+1})f^x\left(x'|n, x_{nt}, \bar{\omega}^{kc}_t\right) + e^{\bar{V}}(n, x_{nt}, \bar{\omega}^{kc}_t, \bar{\omega}^{kc}_{t+1}), \tag{A2}$$

with

$$e^{\bar{V}}\left(n, x_t, \bar{\omega}^{kc}_t, \bar{\omega}^{kc}_{t+1}\right) \equiv \sum_{x'} e^{\bar{V}}\left(x', \bar{\omega}^{kc}_t, \bar{\omega}^{kc}_{t+1}\right) f^x\left(x'|n, x_{nt}, \bar{\omega}^{kc}_t\right).$$

The use of these realized continuation values permits minimal assumptions about households' beliefs over the evolution of the city-specific state variables. Imputing our expression for households' expected continuation values conditional on their own household-level state variables in A2 to our expression for the difference in conditional choice probabilities in A1 yields

$$\ln\left(\frac{p^k_{xnt}}{p^k_{xn't}}\right) = \bar{u}^k_{xnt} - \bar{u}^k_{xn't}$$

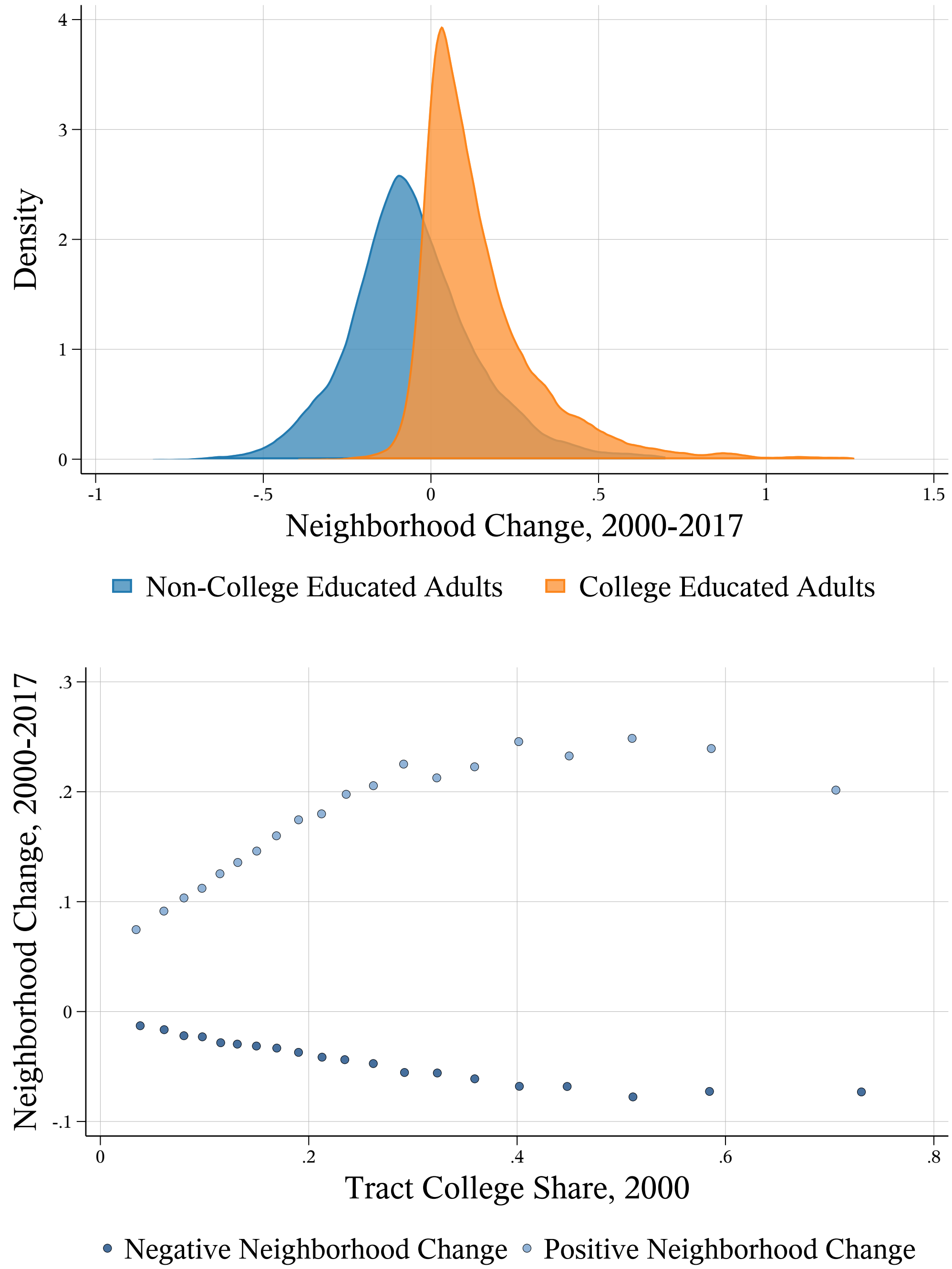

FIGURE A1. Neighborhood Change, 2000 - 2017

These figures are constructed using 2010 delineated census tract population counts for the 100 largest (ranked by population) US CBSAs from the publicly available 2000 NHGIS Census and the American Community Survey 2015-2019 5-year aggregates (Manson et al. 2022). We exclude tracts with fewer than 1,000 adult residents in 2000. Neighborhood change for non-college educated and for college-educated households is defined analogously to equation X. The top panel plots kernel densities of neighborhood change between 2000-2017 among Census tracts that contain the 20% of the CBSA's population that is closest to its CBD. The bottom panel plots our tract-level measure of neighborhood change (y-axis) against the tract's initial share of college-educated workers in 2000, separately for tracts that experienced positive and negative changes in their share of college-educated households.

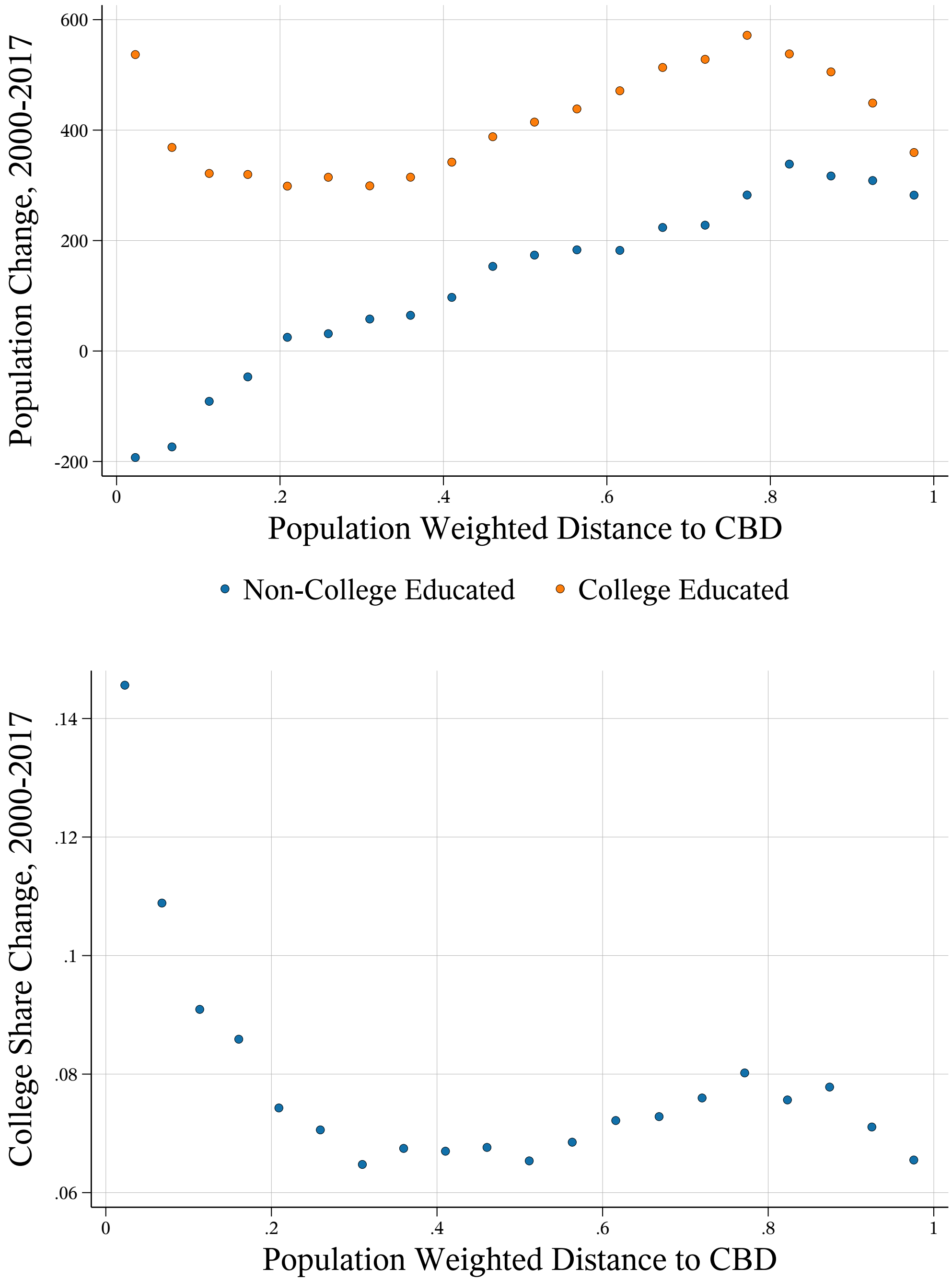

FIGURE A2. Neighborhood Change and Initial College Share

The figure is constructed using 2010 census tract population counts for the 100 largest (ranked by population size) US metropolitan divisions from the publicly available 2000 NHGIS Census and the American Community Survey 2015-2019 5-year aggregates (Manson et al. 2022). We exclude tracts with fewer than 1,000 adult residents in 2000. Neighborhood change is defined by equation X. The top panel plots the degree of population change for college and non-college educated households between 2000-2017 by the population-weighted distance to each metropolitan division's respective CBD. The bottom panel plots the change in the share of college educated households, also by the population-weighted distance to each metropolitan division's respective CBD.

$$+ \delta \left(\sum_{x'} \bar{V}(x', \bar{\omega}^{kc}_{t+1})\, f^{x}\left(x' | n, x_t, \bar{\omega}^{kc}_{t}\right) - \sum_{x'} \bar{V}(x', \bar{\omega}^{kc}_{t+1})\, f^{x}\left(x' | n', x_t, \bar{\omega}^{kc}_{t}\right) \right)$$

$$+ \delta \cdot \tilde{e}(x_t, \bar{\omega}_t^{kc}, \bar{\omega}_{t+1}^{kc})$$

where $\tilde{e}(x_t, \bar{\omega}_t^{kc}, \bar{\omega}_{t+1}^{kc})$ is the difference between the expectational errors when residing in neighborhood n relative to neighborhood n' in period $t+1$,

$$\tilde{e}(x_t, \bar{\omega}_t^{kc}, \bar{\omega}_{t+1}^{kc}) \equiv e^{\bar{V}}(n, x_t, \bar{\omega}_t^{kc}, \bar{\omega}_{t+1}^{kc}) - e^{\bar{V}}(n', x_t, \bar{\omega}_t^{kc}, \bar{\omega}_{t+1}^{kc}).$$

Next, using equation 7 to substitute in for $\bar{V}(x', \bar{\omega}_{t+1}^{kc})$, we obtain,

$$\begin{aligned}
&\ln\left(\frac{p_{xnt}^k}{p_{xn't}^k}\right) - \delta\left(\sum_{x'} \ln\left(p_{x\tilde{n}t}^k\right) f^x\left(x'|n, x_t, \bar{\omega}_t^{kc}\right) - \sum_{x'} \ln\left(p_{x\tilde{n}t}^k\right) f^x\left(x'|n', x_t, \bar{\omega}_t^{kc}\right)\right) \\
&= \bar{u}_{xnt}^k - \bar{u}_{xn't}^k + \delta\left(\sum_{x'} v_{\tilde{n}}^k\left(x', \bar{\omega}_t^{kc}\right) f^x\left(x'|n, x_t, \bar{\omega}_t^{kc}\right) - \sum_{x'} v_{\tilde{n}}^k\left(x', \bar{\omega}_t^{kc}\right) f^x\left(x'|n', x_t, \bar{\omega}_t^{kc}\right)\right) \\
&+ \delta \cdot \tilde{e}(x_t, \bar{\omega}_t^{kc}, \bar{\omega}_{t+1}^{kc})
\end{aligned}$$

Recall that $\tilde{n}$ is a renewal action for both households. Therefore, the household-level state variables are set to the same values for both households regardless of their values in period t. This yields identical continuation values in period $t+1$ for both households. The above expression therefore simplifies to,

$$\begin{aligned}
&\ln\left(\frac{p_{xnt}^k}{p_{xn't}^k}\right) - \delta\left(\sum_{x'} \ln\left(p_{x\tilde{n}t}^k\right) f^x\left(x'|n, x_t, \bar{\omega}_t^{kc}\right) - \sum_{x'} \ln\left(p_{x\tilde{n}t}^k\right) f^x\left(x'|n', x_t, \bar{\omega}_t^{kc}\right)\right) \\
&= \bar{u}_{xnt}^k - \bar{u}_{xn't}^k + \delta\left(MC_t^k(\tilde{n}, n) - MC_t^k(\tilde{n}, n') + \tilde{e}(x_t, \bar{\omega}_t^{kc}, \bar{\omega}_{t+1}^{kc})\right)
\end{aligned}$$

Choosing n' as the city's outside option, applying assumption (b), and substituting in for the neighborhoods' flow utilities provides an equation linear in our model parameters,

$$\begin{aligned}
&\ln\left(\frac{p_{xnt}^k}{p_{xn't}^k}\right) - \delta\left(\sum_{x'} \ln\left(p_{x\tilde{n}t}^k\right) f^x\left(x'|n, x_t, \bar{\omega}_t^{kc}\right) - \sum_{x'} \ln\left(p_{x\tilde{n}t}^k\right) f^x\left(x'|n', x_t, \bar{\omega}_t^{kc}\right)\right) \\
&\quad = \tilde{\alpha}_n^k + \alpha_t^k + \beta_w^k \ln(\bar{I}_{n,t}) - \beta_r^k \ln(r_{n,t}) + \beta_A^k \ln\left(\frac{Coll_{nt}}{Pop_{nt}}\right) \\
&\quad + \beta_\tau^k\left(\sum_{x'} \ln(\tau_{xt}(x')) f^x\left(x'|n, x_{xt}, \bar{\omega}_t^{kc}\right) - \sum_{x'} \ln(\tau_{xt}(x')) f^x\left(x'|n', x_t, \bar{\omega}_t^{kc}\right)\right) \\
&\quad - MC_t^k(n, n_{t-1}) + MC_t^k(n', n_{t-1}) + \delta\left(MC_t^k(\tilde{n}, n) - MC_t^k(\tilde{n}, n')\right) \\
&\quad + \tilde{e}(x_t, \bar{\omega}_t^{kc}, \bar{\omega}_{t+1}^{kc}) + \tilde{\xi}_{nt}^k,
\end{aligned}$$

where $\tilde{\alpha}_n^k = \alpha_n^k - \alpha^{ck}$. To condense notation, we write this equation as

$$Y_{xnn'\tilde{n}t}^k = \tilde{\alpha}_n^k + \alpha_t^k + \beta_w^k \ln(\bar{I}_{n,t}) - \beta_r^k \ln(r_{n,t}) + \beta_A^k \ln\left(\frac{Coll_{nt}}{Pop_{nt}}\right) + \beta_\tau^k \tilde{\tau}_x - \widetilde{MC}_t^k + \nu_{xnn'\tilde{n}t}^k,$$

where

$$\begin{aligned}
Y_{xnn'\tilde{n}t}^k &\equiv \ln\left(\frac{p_{xnt}^k}{p_{xn't}^k}\right) - \delta\left(\sum_{x'} \ln\left(p_{x\tilde{n}t}^k\right) f^x\left(x'|n, x_t, \bar{\omega}_t^{kc}\right) - \sum_{x'} \ln\left(p_{x\tilde{n}t}^k\right) f^x\left(x'|n', x_t, \bar{\omega}_t^{kc}\right)\right) \\
\tilde{\tau}_x &\equiv \sum_{x'} \ln(\tau_{xt}(x')) f^x\left(x'|n, x_t, \bar{\omega}_t^{kc}\right) - \sum_{x'} \ln(\tau_{xt}(x')) f^x\left(x'|n', x_t, \bar{\omega}_t^{kc}\right) \\
\widetilde{MC}_t^k &\equiv MC_t^k(n, n_{xt-1}) - MC_t^k(n', n_{xt-1}) - \delta\left(MC_t^k(\tilde{n}, n) - MC_t^k(\tilde{n}, n')\right) \\
\nu_{xnn'\tilde{n}t}^k &\equiv \tilde{e}(x_t, \bar{\omega}_t^{kc}, \bar{\omega}_{t+1}^{kc}) + \tilde{\xi}_{nt}^k.
\end{aligned}$$

This is the same equation reported in 9.

B.2. First-Step Conditional Choice Probabilities

To see how equation 10 approximates the neighborhood choice problem of households in our dynamic model, start by considering our specification for a household's conditional value function defined in 5,

$$\begin{aligned}
v_n^k\left(x_{it}, \bar{\omega}_t^{kc}\right) &\equiv \bar{u}_n^k\left(x_{it}, \bar{\omega}_t^{kc}\right) + \delta\mathbb{E}_t\left[\bar{V}^k\left(x_{it+1}, \bar{\omega}_{t+1}^{kc}\right)|n, x_{it}, \bar{\omega}_t^{kc}\right] \\
&= \alpha_n^k + \alpha_t^{kc} + \beta_w^k \ln\left(\bar{I}_{n,t}\right) - \beta_r^k \ln\left(r_{n,t}\right) + \beta_A^k \ln\left(\frac{Coll_{nt}}{Pop_{nt}}\right) - MC_t^k(n_t, n_{it-1}) + \tilde{\xi}_{nt}^k \\
&\quad + \sum_{\bar{\tau}=1,2}\left(\delta\mathbb{E}_t\left[\bar{V}^k\left(x_{it+1}(n_t, \bar{\tau}), \bar{\omega}_{t+1}^{kc}\right)|n, x(n_{t-1}, \bar{\tau}_{t-1}), \bar{\omega}_t^{kc}\right]\right. \\
&\quad \left. + \beta_\tau^k \ln(\bar{\tau})\right) \cdot f^{\bar{\tau}}(\bar{\tau}|n_t, x(n_{t-1}, \bar{\tau}_{t-1}), \bar{\omega}^{kc}).
\end{aligned} \tag{A3}$$

Assume for now that $\bar{\tau}_{t-1} = 1$. Then, we can re-write A3 for some neighborhood choice n as,

$$v_n^k\left(x_{it}, \bar{\omega}_t^{kc}\right) = \gamma_{nt}^k + \mu_{\bar{\tau}} \cdot \mathbb{1}\{n = n_{t-1})\} - MC_t^k(n, n_{t-1}) \tag{A4}$$

where,

$$\begin{aligned}
\gamma_{nt}^k = \alpha_n^k + \alpha_t^{kc} + \beta_w^k \ln\left(\bar{I}_{n,t}\right) - \beta_r^k \ln\left(r_{n,t}\right) + \beta_A^k \ln\left(\frac{Coll_{nt}}{Pop_{nt}}\right) + \tilde{\xi}_{nt}^k \\
+ \delta\,\mathbb{E}_t\left[\bar{V}^k\left(x_{it+1}(n_t, \bar{\tau} = 1), \bar{\omega}_{t+1}^{kc}\right)|n, x(n_{t-1}, 1), \bar{\omega}_t^{kc}\right]
\end{aligned}$$

$$\mu_{\bar{\tau}} = f^{\bar{\tau}}(2|n_t, x(n_{t-1},1), \bar{\omega}^{kc}) \Bigg[\beta^k_{\tau} \ln(2) \\ - \delta \cdot \underbrace{\Bigg[\mathbb{E}_t \left[\bar{V}^k \left(x_{it+1}(n_t, \bar{\tau}=1), \bar{\omega}^{kc}_{t+1} \right) | n_t = n_{t-1}, x(n_{t-1},1), \bar{\omega}^{kc}_t \right] \\ - \mathbb{E}_t \left[\bar{V}^k \left(x_{it+1}(n_t, \bar{\tau}=2), \bar{\omega}^{kc}_{t+1} \right) | n_t = n_{t-1}, x(n_{t-1},1), \bar{\omega}^{kc}_t \right] \Bigg] \Bigg]}_{\text{Difference in continuation values w.r.t. } \bar{\tau}_{t-1}}$$

and where $MC^k_t(n, n_{t-1})$ is defined as usual. Equation A4 shows that if the difference in conditional values across tenure status (τ_{t-1}) from staying in one's origin neighborhood was independent of one's origin neighborhood, n_{t-1}, then we could simply estimate our first-step multinomial choice model using A4.[50] It is, however, unlikely that that the difference in continuation values with respect to neighborhood tenure is independent of one's neighborhood origin. To understand why, consider two neighborhoods n and n' in the same CBSA. If neighborhood n provides lower utility than neighborhood n', then *all* residents of neighborhood n in period t are less likely to remain there in period $t+1$ relative to residents in neighborhood n'; when all residents are unlikely to stay in a given neighborhood, the difference in continuation values for incumbent residents with different residential tenures will be small. The converse is true for neighborhoods offering higher utilities to residents.

Equation 10 must therefore capture how the difference in continuation values across neighborhood tenures vary with neighborhoods' mean utilities. We therefore augment equation A4 by incorporating an interaction term between the value of residential tenure and time-varying neighborhood utilities:

$$v^k_n \left(x_{it}, \bar{\omega}^{kc}_t \right) = \gamma^k_{nt} + \mu_{\bar{\tau}} \cdot \mathbb{1}\{n = n_{t-1})\} + \gamma^k_{nt} + \lambda_{\bar{\tau}} \cdot \mathbb{1}\{n = n_{t-1})\} - MC^k_t(n, n_{t-1}).$$

This is the expression that appears in equation 10 and we believe approximates the data generating process implied by our dynamic model well.

B.3. Microfounding Job Market Access and Expected Income

Setup. The workplace choice problem for college graduates is identical to the problem for each type-k household, as reported in section 4.2. In this section we therefore simply extend the index k to also include college graduates. For now, we assume that the labor market for each

[50]When $\bar{\tau}_{t-1} = 2$, the form of γ^k_{nt} remains the same, but $\mu_{\bar{\tau}}$ is no longer scaled by $f^{\bar{\tau}}(2|n_t, x(n_{t-1},1), \bar{\omega}^{kc})$. The current argument therefore remains unchanged when considering $\bar{\tau}_{t-1} = 2$.

k-type is segmented.[51] Recall that in section 4.2, each household's workplace choice problem is to choose which neighborhood to work in to maximize their income net of commute costs:

$$\bar{I}^k_{n,t} \equiv b^k_t \cdot \max_m \frac{z^i_{m,t}}{d_{n,m}} w_{mt},$$
$$= \max_m \frac{z^i_{m,t}}{d_{n,m}} w^k_{m,t},$$

To construct our JMA instrument, we simply amend this workplace choice problem by differentiating wages across industries, I, so that the household's conditional workplace choice problem becomes,

$$\bar{I}^k_{n,t} \equiv \max_{m,I} \frac{\zeta^k_{t,I} z^i_{m,t}}{d_{h,m}} w^k_{m,t},$$
$$= \max_{m,I} \frac{z^i_{m,t}}{d_{h,m}} w^k_{m,t,I},$$

where $z^i_{m,t,I}$ is distributed iid Frechet, $F(z^i_{m,t,I}) = \exp\left(-\left(z^i_{m,t,I}\right)^{-\epsilon^k_c}\right)$ with $\epsilon^k_c > 1$, for each workplace-tract-industry in the city and for each worker i of education group k. $\zeta^k_{t,I}$ is an industry- and type-specific productivity shock that captures the comparative advantage of workers of each type across industries. The probability of worker i of education group k living in tract n taking a job in tract m is then given by,

$$\pi^k_{m|n,t} = \frac{\sum_I (w^k_{m,t,I}/d^k_{n,m})^{\epsilon^k_c}}{\sum_I \sum_{m'} (w^k_{m',t,I}/d^k_{n,m'})^{\epsilon^k_c}},$$
$$= \frac{\sum_I (w^k_{m,t,I}/d^k_{n,m})^{\epsilon^k_c}}{RMA^k_{n,t}},$$

where $RMA^k_{n,t} \equiv \sum_I RMA^k_{n,t,I} \equiv \sum_I \sum_{m'} (w^k_{m',t,I}/d^k_{h,m'})^{\epsilon^k_c}$.

Define labor supply to tract m in time t by $\ell^k_{m,t} = \sum_I \left[\left(w^k_{m,I}\right)^{\epsilon^k_c}\right] FMA^k_{m,t}$, where $FMA^k_{m,t}$, represents the access firms in tract m have to k-type workers. Equating labor supply of k-type workers

[51] Baum-Snow, Hartley, and Lee (2019) show how to extend the model to an integrated labor market with multiple types. The resulting expressions are identical to those we derive here when assuming segmented markets. The derivations in this section follow an established literature microfounding measures of "market access" in economic fundamentals (Donaldson and Hornbeck (2016), Tsivanidis (2022), Baum-Snow and Han (2023)).

to tract m in period t to workers' choice probabilities yields an expression for *FMA* in terms of *RMA*:

$$
\begin{aligned}
\ell^k_{m,t} &= \sum_n \pi^k_{m|n,t} \cdot \pi^k_{n,t} \cdot N^k \\
&= \sum_n \frac{\sum_I (w^k_{m,t,I}/d^k_{n,m})^{\epsilon^k_c}}{RMA^k_{n,t}} \cdot \pi^k_{n,t} \cdot N^k \\
&= \sum_I \left[\left(w^k_{m,t,I} \right)^{\epsilon^k_c} \right] N^k \sum_n \frac{\left(\left(\pi^k_{n,t} \right)^{1/\epsilon^k_c} / d^k_{h,m} \right)^{\epsilon^k_c}}{RMA^k_{n,t}} \\
&\equiv \sum_I \left[\left(w^k_{m,t,I} \right)^{\epsilon^k_c} \right] FMA^k_{m,t}
\end{aligned}
\tag{A5}
$$

The penultimate equality obtains,

$$
FMA^k_{m,t} = N^k \sum_n \frac{\left(\left(\pi^k_{n,t} \right)^{1/\epsilon^k_c} / d^k_{h,m} \right)^{\epsilon^k_c}}{RMA^k_{n,t}}.
$$

Furthermore, dividing both sides of A5 by $\left(d^k_{h,m} \right)^{\epsilon^k_c}$ and summing over m yields an expression for $RMA^k_{n,t}$ in terms of $FMA^k_{m,t}$. We subsequently obtain the following system of equations for *RMA* and *FMA*:

$$
\begin{aligned}
FMA^k_{m,t} &= N^k \sum_n \frac{e^{-\kappa \epsilon^k_c \tau^k_{n,m}} \pi^k_{n,t}}{RMA^k_{n,t}} \\
RMA^k_{n,t} &= \sum_m \frac{e^{-\kappa \epsilon^k_c \tau^k_{n,m}} \ell^k_{m,t}}{FMA^k_{m,t}},
\end{aligned}
\tag{A6}
$$

where we have defined $d^k_{n,m} \equiv e^{\kappa \tau^k_{n,m}}$.[52]

Job Market Access Instrument. To

obtain Given the expressions in A6, one can solve for either

our job market access measure is a linear approximation of $RMA_{n,t}$:

[52] $1 - e^{-\kappa \tau^k_{n,m}}$ represents the portion of time that type-k workers in tract n spend commuting to tract m.

$$JMA_{n,t} = e^{-\kappa\epsilon_c^k \tau_{n,m}^k} \ell_{m,t}^k.$$

We will relate this linear approximation to the literature in our final write-up. Our workplace choice model also implies that expected income discounted by commuting costs for type-k households prior to drawing the vector of neighborhood- and period-specific productivity shocks is given by,

$$\bar{I}_{nt}^k = \Gamma\left(1 - \frac{1}{\epsilon_c^k}\right)\left(RMA_h^k\right)^{1/\epsilon_c^k}, \quad \forall i \in k,$$

where we directly use the identify $RMA_{n,t}^k \equiv \sum_I \sum_{m'} (w_{m',t,I}^k / d_{h,m'}^k)^{\epsilon_c^k}$ in this measure's construction.[53] This is the expression that enters into the consumption associated with households' flow utilities.

Gravity and Forecasting Equations. We use our workplace choice model to motivate estimates of $\kappa\epsilon_c^k$. We follow Baum-Snow, Hartley, and Lee (2019) and estimate $\kappa\epsilon_c^k$ separately for each type-k household in each city, c. Estimating $\kappa\epsilon_c^k$ separately for each type-k household in each city will increase the accuracy of our $RMA_{n,t}^k$ measures and the power of our $JMA_{n,t}$ instruments to the extent that employment growth impacts market access in neighborhoods accessible by longer commute times more in cities in which ϵ_c^k is lower[54]. To obtain our estimates of $\eta^{kc} \equiv \kappa\epsilon_c^k$ we take the log of $\pi_{m|n,t}^k$ to obtain the following gravity equation,

$$\begin{aligned}
\ln(\pi_{m|n,t}^k) &= \ln(RMA_{n,t}^k) + \ln\left(\sum_I \left(w_{m,t,I}^k\right)^{\epsilon_c^k}\right) - \kappa\epsilon_c^k\tau_{n,m}^k \\
&= \alpha_{n,t}^k + \rho_{m,t}^k - \underbrace{(\kappa\epsilon_c^k)}_{\eta^{kc}}\tau_{n,m}^k,
\end{aligned}$$

which we estimate separately for each city and type-k households (including separately for college-educated and White households) using 2010 commute flows calculated in our employee-employer linked data. We also obtain estimates of $\tau_{n,m}^k$ using the median tract commute time for workers in the 2005-2009 ACS surveys between tracts n and m for all tract combinations that are reported to have positive commute flows for each k-type worker. Since the ACS only samples a small portion of the population each year, many commute times and flows are not observed

[53] In practice we regress observed workplace wages discounted by commute costs on this measure of $RMA_{n,t}$ and predict $\bar{I}_{nt}^k$ solely using variation in $RMA_{n,t}$. This is to account for the fact that larger labor markets have mechanically have higher levels of $RMA_{n,t}^k$. We will include all details on this construction once we write up the empirical portion of the paper.

[54] This explanation is similar to the one given in Baum-Snow, Hartley, and Lee (2019).

in our data. To estimate the remaining commute times, we follow Baum-Snow, Hartley, and Lee (2019) and construct an empirical forecasting model to predict commute times between all neighborhood pairs using the distance between neighborhood centroids and the corresponding city's CBD,

$$\ln \tau^k_{n,m} = \alpha^k_d \ln \text{Distance}_{n,m} + \alpha^k_r \ln(\text{Home CBD Dis})_n + \alpha^k_w \ln(\text{Work CBD Dis})_m + \nu_c + u^k_{h,m,c}.$$

As implied by the notation, we estimate these equations separately for each type-k worker.

B.4. Simulating Experienced Welfare

This subsection derives our neighborhood- and period-specific measures of expected welfare for our target population of low-income incumbent renters. Upon forecasting our census-tract aggregates to 2030, we solve for neighborhood-level expected welfare via backward induction under the assumption that the residential choices of our target population is in steady state.[55] We then proceed as follows:[56]

(a) Assume steady state in 2030 so that $s_{n,\bar{\tau},2031} = s_{n,\bar{\tau},2030} \forall n, \bar{\tau}$. Then, solve for the unobserved portions of neighborhood utility that induces this steady-state equilibrium,

$$
\begin{aligned}
s_{n,\bar{\tau}=2,2030} =& \frac{\exp\left(\tilde{\nu}(n,n') + \xi^k_{n,t} + \delta\mathbb{E}_t\left[V^k(s_{i,t+1})|n, s_{it}(n,\bar{\tau}=2)\right]\right)}{\exp(MCOO) + \sum_{\tilde{n}} \exp\left(\tilde{\nu}(\tilde{n},n') + \xi^k_{\tilde{n},t} + \delta\mathbb{E}_t\left[V^k(s_{i,t+1})|\tilde{n}, s_{it}(n,\bar{\tau}=2)\right]\right)} \\
& \cdot \left(s_{n',\bar{\tau}=2,2030} + s_{n',\bar{\tau}=1,2030}\mathbb{P}(\bar{\tau}=2|\bar{\tau}=1, n'=n)\right) \\
s_{n,\bar{\tau}=1,2030} =& \sum_{n'\neq n}\left[\frac{\exp\left(\tilde{\nu}(n,n') + \xi^k_{n,t} + \delta\mathbb{E}_t\left[V^k(s_{i,t+1})|n, s_{it}(n,\bar{\tau}=2)\right]\right)}{\exp(MCOO) + \sum_{\tilde{n}} \exp\left(\tilde{\nu}(\tilde{n},n') + \xi^k_{\tilde{n},t} + \delta\mathbb{E}_t\left[V^k(s_{i,t+1})|\tilde{n}, s_{it}(n,\bar{\tau}=2)\right]\right)} \cdot s_{n',\bar{\tau}=2,2030}\right] \\
& + \sum_{n'\neq n}\left[\frac{\exp\left(\tilde{\nu}(n,n') + \xi^k_{n,t} + \delta\mathbb{E}_t\left[V^k(s_{i,t+1})|n, s_{it}(n,\bar{\tau}=1)\right]\right)}{\exp(MCOO) + \sum_{\tilde{n}} \exp\left(\tilde{\nu}(\tilde{n},n') + \xi^k_{\tilde{n},t} + \delta\mathbb{E}_t\left[V^k(s_{i,t+1})|\tilde{n}, s_{it}(n,\bar{\tau}=1)\right]\right)} \cdot s_{n',\bar{\tau}=1,2030}\right] \\
& + \frac{\exp\left(\tilde{\nu}(n,n') + \xi^k_{n,t} + \delta\mathbb{E}_t\left[V^k(s_{i,t+1})|n, s_{it}(n,\bar{\tau}=1)\right]\right)}{\exp(MCOO) + \sum_{\tilde{n}} \exp\left(\tilde{\nu}(\tilde{n},n') + \xi^k_{\tilde{n},t} + \delta\mathbb{E}_t\left[V^k(s_{i,t+1})|\tilde{n}, s_{it}(n,\bar{\tau}=1)\right]\right)} \\
& \cdot \left(s_{n',\bar{\tau}=1,2030}(1 - \mathbb{P}(\bar{\tau}=2|\bar{\tau}=1, n'=n))\right)
\end{aligned}
$$

[55] The current version of the simulations assume renters have perfect foresight, but extending the simulations to assume renters have rational expectations is straightforward.

[56] Note in the main text that the reason we can't deduce welfare from any time in the analysis period is because households' initial conditions prevent us from inferring welfare from their choices.

where $\tilde{\nu}(n, n') \equiv \beta_A^k \ln\left(Col\,l\,Share_{n,t}\right) - \beta_C^k \ln(HC_{n,2030}^k) - MC_t^k(n, n')$.

We have $2 \times n$ unknowns of the form $\xi_{n,t}^k + \delta\mathbb{E}\left[V^k(s_{i,t+1})|n, s_{it}(n, \bar{\tau} = 2)\right]$ and $2 \times n$ equations. We can therefore solve for the clusters $\xi_{n,t}^k + \delta\mathbb{E}\left[V^k(s_{i,t+1})|n, s_{it}(n, \bar{\tau} = 2)\right]$.

(b) Given the estimates of neighborhood mean utility and steady-state continuation values, we next solve for the neighborhood mean utilities in 2029 by equating the shares of long-term residents in 2029 with their respective choice probabilities, substituting in for the 2030 values of $\xi_{n,t}^k + \delta\mathbb{E}\left[V^k(s_{i,t+1})|n, s_{it}(n, \bar{\tau} = 2)\right]$:

$$s_{n,\bar{\tau}=2,2029} = \frac{\exp\left(\tilde{\nu}(n, n') + \xi_{n,t}^k + \delta\left(\log\left(\sum_n \exp(\tilde{\nu} + \widehat{\xi + EV})\right) + \gamma\right)\right)}{\exp(MCOO) + \sum_{\tilde{n}} \exp\left(\tilde{\nu}(\tilde{n}, n') + \xi_{\tilde{n},t}^k + \delta\left(\log\left(\sum_n \exp(\tilde{\nu} + \widehat{\xi + EV})\right) + \gamma\right)\right)} \cdot \left(s_{n',\bar{\tau}=2,2028} + s_{n',\bar{\tau}=1,2028}\mathbb{P}(\bar{\tau} = 2|\bar{\tau} = 1, n' = n)\right)$$

We now have n equations and n unknowns (the ξs). We solve for the fixed point.

(c) We then compute expected welfare for each neighborhood for each type, and start from step *b* (but with the *xi*'s and the *EV*'s computed separately this time).

Appendix C. Data Appendix

C.1. Data Descriptions

Master Address File Extract. See John Sullivan's and Katie Genadek's paper.

Master Address Auxiliary Reference File. See John Sullivan's and Katie Genadek's paper.

American Community Surveys.

Longitudinal Employer-Households Dynamics.

Longitudinal Business Database.

Core Logic.

C.2. Data

We link persons across the MAF-ARF, LEHD, and ACS with unique person identifiers called protected identity keys (PIKs) which is assigned to individuals across data sets by the Census Bureau via probabilistic linking (Wagner and Layne 2014).

Residential Histories and Housing Characteristics. Individuals' residential histories are obtained using the Auxiliary Reference File associated with the US Census Bureau's Master Address File (MAF-ARF). The MAF-ARF uses data combined from several sources of federal administrative data and internal Census Bureau data to assign yearly locations to around 90% of the entire US population starting in 2000. Individuals' residential locations are recorded up to the precise address and each address is provided with its own unique housing unit identifier. These unique identifiers facilitate links with housing unit characteristics recorded in the Census Bureau's Master Address File Extract (MAF-X), which is itself a complete repository of every residential housing unit used in Census Bureau surveys.[57] Included among these housing unit characteristics is the type of housing unit each address corresponds to (e.g. single family home, trailer, or multi-unit building).

Earnings and Workplace Locations. Data on individuals' employment, earnings, and workplace locations are obtained from the Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) files. The LEHD data are derived from states' unemployment insurance agencies. The Census Bureau standardizes these data across states and supplements them with additional information on individuals and their employers (Abowd et al. 2009). The LEHD covers roughly 95% of private sector workers as well as state and local government employees. It does, however, exclude Federal employees, the self-employed, and armed service members. We use this data to construct annual measures of individuals' earnings and workplace locations for all workers across the 28 states for which we have data access.

Sociodemographic Characteristics. Individuals' sociodemographic characteristics are obtained from both the LEHD and the ACS. Individuals in the LEHD are provided with basic sociodemographic information obtained from the Decennial Censuses and the ACS. We use the race, ethnicity, sex, date-of-birth, and education variables recorded in the LEHD.[58] Where possible, we supplement the sociodemographic information present in the LEHD with additional

[57] In practice, the MAF-ARF is constructed using the MAF-X as a source file. We document the construction of these data sources in appendix C.1.

[58] As we outline in appendix C.1, we are careful to distinguish when these variables are imputed versus when they are observed in the Decennial Censuses and ACS.

individual- and household-level characteristics from the 2005-2021 ACS, such as parental status, housing tenure, and self-reported property values.

Business Establishment Data. Business establishment data come from the Census Bureau's Longitudinal Business Database (LBD). Obtained from Census Bureau business surveys and business tax filings, the LBD records annual geocoded establishment-level data on total employment, total payroll, and 6-digit NAICS industry code (Chow et al. 2021). Each establishment has a unique identifier that facilitates linking establishments over time. We primarily use the LBD, however, to construct tract-level aggregates of industry employment, which we in turn use to form our hyper-local Bartik shift-share instrument.

C.3. Sample Construction

Reduced-form and Welfare Analysis Samples. We construct two individual-level panels for our reduced-form analysis and welfare calculations, one from 2002-2019 and another from 2010-2019. These panels use as their basis the residential history panel constructed from the MAF-ARF. We merge to these residential history panels individuals' earnings and workplace locations from the LEHD as well as additional sociodemographic characteristics from the ACS. These merges are facilitated by a unique person identifier called a protected identity key (PIK) which is assigned to individuals across data sets by the Census Bureau via probabilistic linking (Wagner and Layne 2014). We aggregate earnings to the housing unit level and call these housing units "households". We assign the highest earner of each household in 2002 (2010) as the household head, and subsequently focus on the residential location choices of these household heads.[59]

We restrict our sample to household heads that are between 25 and 65 years of age and who have positive earnings in the LEHD in 2002 (2010). To focus our analysis on low-income residents, we restrict our sample to household heads earning in the bottom tercile of their respective core-based statistical area (CBSA) and decadal age band in 2002 (2010), and drop household heads ever living in a household with annual earnings greater than $150,000 (2010 dollars).[60] We finally restrict our analysis to household heads residing in the urban cores of the 100 largest CBSAs that are present in the 28 states for which we can access data in the LEHD. We define urban cores using a similar method to Hwang and Lin (2016) and Couture and Handbury (2023), labeling them as the set of census tracts associated with each CBSA that contain the 50% of the

[59]We detail how we handle changes to household formations, out-of-sample migrations, deaths, and missing observations separately for each analysis sample in appendix C.1.

[60]CBSAs consist of counties associated with an urban core of at least 10,000 persons as well as adjacent counties that are deemed integrated through commuting ties.

CBSA's population that is closest to its central business district (CBD).[61] Finally, for many of our analyses using our 2010-2019 sample, we exclude residents who moved into their 2010 census tract between 2005-2009. Summary statistics for our 2010-2019 sample are presented in table **??**, while summary statistics for our 2002-2019 sample are presented in appendix table.

Structural Estimation Sample. Our structural estimation sample is constructed using individual-level data from 2010-2019. The sample is defined identically to the reduced-form and welfare analysis samples aside from the following deviations. First, to maximize sample size we don't require individuals to be present in the sample for the baseline year (i.e. 2010); we only require that individuals are present in the sample for at least two consecutive years. Second, household head status is determined in the second year the individual is observed with positive earnings in the LEHD. This is to account for the fact that first-year earnings in the LEHD are less likely to correspond to a full year's income. Third, to reduce computational burdens, in addition to only focusing on the years 2010-2019, we further restrict our analysis sample to the largest 50 CBSAs present in our LEHD data.[62] Summary statistics for our structural sample are in the appendix.

C.4. Detailed Sample Construction

C.5. Aggregate Variable Construction

i.e. how do we calculate the annual measures of neighborhood change etc.

Appendix D. Reduced-Form Appendix

D.1. Control Variables

Household-Level Controls. Our vector of household-level controls, X_i, is designed to capture household characteristics that jointly determine one's origin location and our set of outcome

[61] Our CBD definitions come from Fee and Hartley (2013). These CBD definitions are with respect to 2008 CBSA delineations. To keep consistent with these CBD definitions, we therefore use 2008 delineations of CBSAs throughout our analysis. Moreover, while each CBSA is associated with a primary urban center, some CBSAs additionally contain secondary urban centers called metropolitan divisions that have their own CBD. Although our low-income cutoff is constructed using the earnings distribution of the entire CBSA, our urban core cutoffs are particular to each urban center's population, including metropolitan divisions. We believe these choices best capture our target population of low-income urban residents.

[62] For this structural sample, we additionally only focus on the primary urban cores associated with each CBSA. Constructing monocentric cities facilitates an intuitive outside option for each CBSA.

TABLE A1. Sample Characteristics: Household Heads in 2010

		Renters		Owners	
	Total	Non-Black	Black	Non-Black	Black
Panel A: Household Head Characteristics, 2010					
Household Income	-108.4***	-91.22***	-49.51	21.85	h
	(-11.60)	(-10.34)	(-0.57)	(0.29)	h
Commute Distance	-108.4***	-91.22***	-49.51	21.85	h
	(-11.60)	(-10.34)	(-0.57)	(0.29)	h
College Degree	-108.4***	-91.22***	-49.51	21.85	h
	(-11.60)	(-10.34)	(-0.57)	(0.29)	h
Immigrant	-108.4***	-91.22***	-49.51	21.85	h
	(-11.60)	(-10.34)	(-0.57)	(0.29)	h
Age	-108.4***	-91.22***	-49.51	21.85	h
	(-11.60)	(-10.34)	(-0.57)	(0.29)	h
Female	-108.4***	-91.22***	-49.51	21.85	h
	(-11.60)	(-10.34)	(-0.57)	(0.29)	h
Household Size	-108.4***	-91.22***	-49.51	21.85	h
	(-11.60)	(-10.34)	(-0.57)	(0.29)	h
Parent	-108.4***	-91.22***	-49.51	21.85	h
	(-11.60)	(-10.34)	(-0.57)	(0.29)	h
Panel B: Household Heads' Tract Characteristics, 2010					
Median Rents	-108.4***	-91.22***	-49.51	21.85	h
	(-11.60)	(-10.34)	(-0.57)	(0.29)	h
Median Property Value	-108.4***	-91.22***	-49.51	21.85	h
	(-11.60)	(-10.34)	(-0.57)	(0.29)	h
Share White	-108.4***	-91.22***	-49.51	21.85	h
	(-11.60)	(-10.34)	(-0.57)	(0.29)	h
Share College-Educated	-108.4***	-91.22***	-49.51	21.85	h
	(-11.60)	(-10.34)	(-0.57)	(0.29)	h
Share College-Educated and White	-108.4***	-91.22***	-49.51	21.85	h
	(-11.60)	(-10.34)	(-0.57)	(0.29)	h
Population	-108.4***	-91.22***	-49.51	21.85	h
	(-11.60)	(-10.34)	(-0.57)	(0.29)	h
Distance to CBD	-108.4***	-91.22***	-49.51	21.85	h
	(-11.60)	(-10.34)	(-0.57)	(0.29)	h
Observations	74	74	74	74	n

Notes: Table reports mean characteristics with standard errors in parentheses. Sample consists of household heads in the 2010-2019 reduced-form and welfare analysis panel. Panel A reports household head characteristics. Panel B reports characteristics of household heads' census tracts. Dollars are deflated to 2010, and Census tracts are delineated by 2010 boundaries. Sources: 2005-2021 ACS, LEHD, CoreLogic, and MAF-ARF. Details on construction of tract aggregates are in the appendix. The Parent variable is calculated only for household heads present in the ACS 2005-2021, and is inferred by the reported age of the child in the year they are survey respondents. The college degree variable is only computed for PIKs for whose education variables are not imputed or for whom we could link to our ACS surveys.

variables, Δy_i and $h(t|i)$. These controls include i) average household income in defined as the mean income of all adults residing at the same address during 2010; ii) household size defined as the number of adults residing at the same address;[63] iii) a second-order polynomial in the household head's date of birth; iv) indicators for the household head's race, immigrant status, sex, and college degree attainment;[64] and v) the length of the household head's prior residential tenure in neighborhood $n(i)$. All variables are defined in our base year, 2010.

Neighborhood-Level Controls. Our vector of neighborhood-level controls, $X_{n(i)}$, is designed to capture characteristics of the household's origin neighborhood that are potentially correlated with our measure of neighborhood change and our set of outcome variables, Δy_i and $h(t|i)$. We first include neighborhood-level controls defined in 2010 which include i) the share of adults in the neighborhood with a college degree; ii) the share of adults in the neighborhood that identify as non-hispanic and white; iii) the median income among working age adults residing in the neighborhood; iv) the median property value and rent payment in the neighborhood; v) the degree of neighborhood churn which we define as the share of adults residing in the neighborhood during 2008 who also remain in the neighborhood during 2009; vi) second-order polynomials in the distance to the metropolitan division's CBD, measured in both the physical distance as well as in the cumulative share of the metropolitan division's residents residing closer to the CBD than those in tract $n(i)$; vii) fixed effects capturing five bands equidistant from the metropolitan division's CBD; and viii) the 5-year lag in our measure of neighborhood change.

In addition to our neighborhood-level controls defined in 2010, we also include a few contemporaneous neighborhood-level controls that are similarly designed to capture *changing* characteristics of the household's origin neighborhood that are potentially correlated with our measure of neighborhood change and our set of outcome variables, Δy_i and $h(t|i)$. In particular, we control for changes in job market access to *tradable* industries among low-skilled workers. In terms of equation 14, these changes in job market access among tradable industries for low-skilled workers are defined as $\Delta \sum_{d\in\mathfrak{T}} JMA_{ndt}\tilde{\theta}^c_d$, where Δ corresponds to the difference in our JMA measure across 2010 and 2019, and $\tilde{\theta}^c_d$ is the share of non-college educated workers employed in industry d in CBSA c's state.[65] We finally include in $X_{n(i)}$ a third-order polynomial in the change in neighborhood $n(i)$'s total population.

[63] We topcode this value to 10, as a few addresses record an infeasibly high number of adult residents.

[64] While we run our analysis separately for Black and non-Black headed households, we include finer racial distinctions as part of our controls.

[65] Note that including non-tradable industries into this measure of job market access would induce a classic "bad controls" problem, as changes in non-tradable employment can easily be conceptualized as an outcome of neighborhood change (Angrist and Pischke 2009).

Recall that all specifications additionally include CBSA-level fixed effects α_{CBSA}, ensuring that variation in gentrification comes from across neighborhoods in the same CBSA.

D.2. Identification

A causal interpretation of our coefficients of interest, β_{NC}^{Cox} and β_{NC}^{LP}, is based on the conditionally random assignment of neighborhood change across neighborhoods between 2010 and 2019. That is, conditional on our control variables, we assume there are no unobserved neighborhood- or household-level characteristics that are correlated with our measure of neighborhood change. Given our setup, to interpret β_{NC}^{Cox} and β_{NC}^{LP} causally, we much ensure two conditions are satisfied:

(a) First, residents in observably similar neighborhoods in 2010 must not differ in unobservable ways that correlate with our outcome variables, Δy_i and $h(t|i)$. If residents moving into gentrifying neighborhoods prior to 2010 are observably similar to their incumbent residents but different in unobservable ways that affect Δy_i and $h(t|i)$ (access to familial wealth, for example), β_{NC}^{Cox} and β_{NC}^{LP} will partly reflect sample selection.

(b) Second, neighborhoods experiencing gentrification throughout our analysis period must not be undergoing changes in their unobserved characteristics which independently predict incumbents' outcomes. While changes to unobserved public and private amenities caused by increased neighborhood demand among college-educated residents constitute part of our treatment, our identifying variation must be purged of shocks to neighborhood characteristics that affect incumbent residents' outcomes independently of gentrification. These unobserved shocks may include changes in tradable job market access (Kain (1962); Miller (2021)), changes in transportation infrastructure that precedes gentrification (LeRoy and Sonstelie (1983); Glaeser, Kahn, and Rappaport (2008); Curci and Yousef (2022)), or changes in neighborhood valuations resulting from secular trends in preferences or from shifts to within-city income distributions (Brueckner (1987); Couture et al. (2023)). That suburbanization has continued unabated throughout our analysis periods - particularly for Black residents - makes these concerns especially acute (Bartik and Mast (2023); Couture and Handbury (2023)).

We take steps to help ensure conditions (a) and (b) are met. First, in addition to our rich household-level controls, we often subset our sample to those who have lived in their origin neighborhood for at least 5 years. While this decision focuses our analysis on longtime renters, it helps disambiguate the outcomes of gentrifiers and the outcomes of our target population of incumbent low-income renters, reducing the potential for sample bias. This is especially true when considering we control for the five-year lag in our measure of gentrification. Indeed, for sample selection to influence our results, it must be that low-income renter households predict with some accuracy how trends in neighborhood change will vary five years in the future and base their current residential choices on these predictions in a way that is uncorrelated with both their own observable characteristics (measured in 2010) and their chosen neighborhoods' observable characteristics (also measured in 2010).

Second, we are careful to control for changes in neighborhoods' unobserved characteristics that may independently predict incumbents' outcomes. By controlling for changes in access to low-skill tradable employment opportunities, we help ensure aggregate economic conditions that may independently affect neighborhood composition are not driving incumbents' outcomes. This concern is relevant given stark evidence of differential mobility responses to aggregate labor demand shocks across low- and high-skill workers (Notowidigdo 2020). By controlling for a cubic in neighborhood-level population change over our analysis period, we further control for unobserved neighborhood-level shocks that similarly affect incumbent residents and potential gentrifiers. To take an extreme example, consider a natural disaster during our analysis period that leads to a large depopulation of affected census tracts. In this scenario, we will observe low rates residential tenure among incumbent residents which we - without our population controls - would falsely attribute to a decline in our measure of neighborhood change. Note that the inclusion of our population controls implies that the coefficients on our measure of neighborhood change should be interpreted as the effect of changing neighborhood *composition* on incumbents' outcomes. Finally, our rich set of geographical controls ensure that we are comparing outcomes for incumbent residents across origin tracts that are equidistant from the Metropolitan Division's CBD, mitigating bias from secular trends toward suburbanization among our target population.

While we ensure our controls are carefully chosen to mitigate the impact of sample selection and changes in unobserved neighborhood-level characteristics, our estimates are robust to the exclusion of any small subset of control variables. It is finally worth noting that we explored using our instrumental variables detailed in section 6 to estimate our reduced-form equations. We find our reduced-form estimates are sensitive to the choice and composition (i.e. years and industries selected) of the instruments, indicating substantial heterogeneity in the complier characteristics of incumbent renters' origin neighborhoods across our instruments. It is therefore difficult to interpret the reduced-form IV estimates without placing more structure on our reduced-form equations to understand how each relevant equilibrium object (e.g. rents, non-tradable job market access, and neighborhood socioeconomic composition) mediates the impact of gentrification on incumbent renters. For now we report our OLS estimates which we believe offer a more transparent depiction of the impact of neighborhood change on incumbent renters' observable outcomes.

Cox Proportional Hazards Assumption. Identification in our Cox Proportional Hazards models further require that the proportional hazards assumption is met. Namely, that the impact of neighborhood change on incumbents' hazard rates are constant across each year between 2010 and 2019. We test this assumption by plotting – log(– log(survival probability)) against log(time) separately for incumbent residents originally residing in neighborhoods within each decile of our measure of neighborhood change. We observe parallel lines across all deciles of neighborhood change, consistent with the proportional hazards assumption. We also test the null hypothesis that the corresponding Schoenfeld residuals for our measure of neighborhood change are not serially correlated.[66] We report the results of Stata's `phtest` command for our measure of neighborhood change in Table A2, which indicates that we cannot reject the null at the 95% confidence level for our full samples of Black and non-Black incumbent renters (though we are close to doing so for non-Black renters). Finally, our linear probability models examining the impact of neighborhood change on the probability households leave their origin neighborhoods during 2010-2019 yield quantitatively similar results, further supporting the proportional hazards assumption.

TABLE A2. Schoenfeld Residuals

Shoenfeld Residuals	Black	Non-Black
$\mathbb{P} \geq \chi^2$	0.317	0.056

Notes: Table A2 reports the results of Stata's `phtest`, testing the null hypothesis that the Schoenfeld residuals are not serially correlated. The test statistic is distributed as X^2 under the null hypothesis of no serial correlation.

[66] The Schoenfeld residuals correspond to the difference between the observed covariate values and the expected covariate values under the Cox proportional hazards model for incumbent renters at each year between 2010 and 2019. If the proportional hazards assumption holds, these residuals should not be serially correlated (Kleinbaum and Klein 1996).

TABLE A3. Cox Model: Effect of Gentrification on Incumbent Renters' Tenure (2010-2019)

Hazard Rate	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Neighborhood Change	-0.0688	-0.107*	-0.0793	0.187*	0.195	0.204*	0.369	0.323*	-0.777**	-0.264	0.165	-0.901*
	(0.0542)	(0.0439)	(0.132)	(0.0781)	(0.147)	(0.871)	(0.204)	(0.108)	(0.286)	(0.325)	(0.555)	(0.411)
Select Household-level Controls												
Household Income ($1,000s)	-0.00381***	-0.00341***	-0.00320***	-0.00277***	-0.00273***	-0.00275***	-0.00326***	-0.00257***	-0.00365***	-0.00300***	-0.00301***	-0.00476***
	(0.000155)	(0.000110)	(0.000413)	(0.000271)	(0.000666)	(0.000343)	(0.000858)	(0.000453)	(0.000526)	(0.000438)	(0.000887)	(0.000733)
Household Head DOB	0.0157***	0.0157***	0.0167***	0.143***	0.0157***	0.0145***	0.0151***	0.0152***	0.0175***	0.0144***	0.0189***	0.0130***
	(0.000216)	(0.000154)	(0.000581)	(0.000381)	(0.000948)	(0.000503)	(0.00124)	(0.000675)	(0.000734)	(0.000593)	(0.00120)	(0.000973)
Residential Tenure	-0.102***	-0.105***	-0.116***	-0.116***	-0.112***	-0.118***	-0.120***	-0.116***	-0.116***	-0.114***	-0.105***	-0.106***
	(0.000833)	(0.000555)	(0.00375)	(0.00230)	(0.00601)	(0.00311)	(0.00782)	(0.00411)	(0.00478)	(0.00343)	(0.00804)	(0.00596)
Select Neighborhood-level Controls												
Neighborhood Churn	-0.328***	-0.358***	-0.901***	-0.698***	-0.696***	-0.693***	-0.740***	-0.662***	-0.928***	-0.612***	-1.145***	-0.712***
	(0.0171)	(0.0117)	(0.102)	(0.0639)	(0.144)	(0.0825)	(0.202)	(0.112)	(0.146)	(0.101)	(0.232)	(0.179)
Rent ($1,000s)	0.272***	0.197***	0.0661	0.122***	0.135*	0.100***	0.146	0.100*	0.0749	0.198**	-0.0528	0.134
	(0.0209)	(0.0143)	(0.0452)	(0.0313)	(0.0617)	(0.0303)	(0.0844)	(0.0399)	(0.0664)	(0.0654)	(0.121)	0.0784
College Share	.141***	0.212***	0.235***	0.285***	-0.0866	0.143***	0.116	0.138*	0.923***	0.818***	0.704*	1.039***
	(0.0275)	(0.0178)	(0.0662)	(0.0350)	(0.108)	(0.0494)	(0.156)	(0.0676)	(0.207)	(0.163)	(0.352)	(0.252)
White Share	0.0975***	-0.113***	0.221***	-0.155***	0.215***	-0.0654	0.0348	-0.0666	0.245***	-0.195***	0.140*	-0.123
	(0.0150)	(0.0133)	(0.0321)	(0.0246)	(0.0578)	(0.0367)	(0.0807)	(0.0461)	(0.0383)	(0.0347)	(0.0586)	(0.0629)
Sample Restrictions												
Race	Black	Non-Black	Black	Non-Black	Black	Non-Black	Black	Non-Black	Black	Non-Black	Black	Non-Black
Longtime Renters			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Initial College Share					High	High	High	High	Low	Low	Low	Low
Fraction Developed							High	High			Low	Low
N (1,000s)	314	688	56	156	21	89	13.5	54.5	35	67	10.5	20.5

Notes: Coefficients correspond to the percent change in the hazard rate from a one unit increase in the corresponding independent variable. A one unit increase in our measure of neighborhood change corresponds to a one hundred percentage point increase. Every specification includes the full set of controls listed and detailed in Appendix D. Standard errors in parentheses are clustered at the origin census tract level. Longtime renters are renters who have resided in their origin Census tract since at least 2005. Tracts with a "High" ("Low") initial college share are tracts whose share of college-educated adults in 2010 is above (below) the population-weighted median in our full sample. Tracts with a "High" ("Low") fraction developed are tracts whose developed land area in 2011 is above (below) the population-weighted median in our full sample.

Table A4. Linear Probability Model: Effect of Gentrification on Incumbent Renters (2010-2019)

Outcome Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Household Outcomes												
Leave Tract	-0.0176	-0.0263	-0.016	0.07*	0.104	0.0668*	0.129	0.111***	-0.305***	-0.0888	-0.0458	-0.303
	(0.0223)	(0.0174)	(0.0509)	(0.0278)	(0.0571)	(0.0308)	(0.0738)	(0.0368)	(0.106)	(0.0894)	(0.254)	(0.158)
Moved > 5 Miles	0.038	0.0167	0.0066	0.0803**	0.0438	0.0709*	0.023	0.0926**	-0.0324	-0.0049	0.386	-0.205
	(0.0217)	(0.0171)	(0.0406)	(0.0261)	(0.0505)	(0.0294)	(0.0649)	(0.0339)	(0.0783)	(0.0802)	(0.223)	(0.154)
Leave CBSA	-0.0519***	-0.0375**	-0.0158	0.0087	-0.0194	0.0191	-0.0200	0.0179	0.0012	-0.0194	-0.0278	-0.1500
	(0.013)	(0.0122)	(0.0238)	(0.0160)	(0.0300)	(0.0178)	(0.0387)	(0.0205)	(0.0451)	(0.0432)	(0.1130)	(0.089)
Income	2,872**	2,268**	-932	-86.05	-1,458	-926.6	-1,875	311.5	-9,138*	-551.8	-23,950**	-3,874
	(948.8)	(736.4)	(2,170)	(1,375)	(2,748)	(1,547)	(3,498)	(1,816)	(3,854)	(3,427)	(9,225)	(6,444)
Commute Distance	-0.2280	-1.905***	0.637	0.5900	3.106	-0.234	3.651	1.0800	-1.854	3.146	-9.659	-3.379
	(0.776)	(0.531)	(1.609)	(1.002)	(2.059)	(1.146)	(2.529)	(1.352)	(3.132)	(2.775)	(6.284)	(5.352)
Panel B: Experienced Tract Characteristics												
Rent	0.271***	0.0960***	0.389***	0.222***	0.159**	0.132***	0.171*	0.153***	0.863***	0.607***	0.759**	0.485***
	(0.0472)	(0.0288)	(0.0620)	(0.0329)	(0.0546)	(0.0335)	(0.0701)	(0.0434)	(0.145)	(0.096)	(0.241)	(0.131)
College Share	0.726***	0.299***	1.079***	0.567***	0.539***	0.392***	0.570***	0.379***	3.726***	4.108***	3.559***	3.897***
	(0.104)	(0.0598)	(0.131)	(0.0788)	(0.0714)	(0.0525)	(0.0845)	(0.0689)	(0.250)	(0.236)	(0.565)	(0.357)
White College Share	-0.298	0.635***	0.277	1.269***	0.822*	0.449***	0.806	0.358***	7.758*	7.567***	2.832	5.206***
	(0.765)	(0.156)	(1.393)	(0.226)	(0.384)	(0.0863)	(0.419)	(0.098)	(3.951)	(1.167)	(7.965)	(1.169)
Sample Restrictions												
Race	Black	Non-Black	Black	Non-Black	Black	Non-Black	Black	Non-Black	Black	Non-Black	Black	Non-Black
Longtime Renters			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Initial College Share					High	High	High	High	Low	Low	Low	Low
Fraction Developed							High	High			Low	Low
N (1,000s)	277	599	50	138	19	78.5	12	48.5	31	59.5	9.3	18

Notes: Coefficients correspond to the impact of a 100 percentage point increase in our measure of gentrification on a change in the associated outcome variable over 2010 to 2019. The dependent variables “Leave Tract”, “Moved > 5 Miles”, and “Leave CBSA” are all indicator variables equal to one if the corresponding condition is satisfied. Income and rents are measured in 2010 dollars and commute distance is measured in miles. Experienced neighborhood characteristics are measured in percent changes. Every specification includes the full set of controls listed and detailed in Appendix D. Standard errors in parentheses are clustered at the origin census tract level. Longtime renters are renters who have resided in their origin Census tract since at least 2005. Tracts with a “High” (“Low”) initial college share are tracts whose share of college-educated adults in 2010 is above (below) the population-weighted median in our full sample. Tracts with a “High” (“Low”) fraction developed are tracts whose developed land area in 2011 is above (below) the population-weighted median in our full sample.