Quantifying the Welfare Effects of Gentrification on Incumbent Low-Income Renters*

Robert French, Loyola Marymount University
Ashvin Gandhi, UCLA Anderson & NBER
Valentine Gilbert, University of Houston

May 1, 2026

Preliminary and incomplete.
Please do not cite or circulate without author permission.

Updated frequently—please click here for the latest draft.

How does gentrification affect the welfare of incumbent residents of low-income neighborhoods? This paper investigates how low-income renters in gentrifying neighborhoods fared relative to renters in neighborhoods in the same metro area that stayed poor during 2010-2019. We link person-level administrative US Census data to construct an annual panel that tracks the earnings, workplaces, and residential addresses of over 3 million US low-income urban renter households. We use these data to conduct a comprehensive descriptive analysis and to estimate a dynamic structural model of residential and workplace choice. We find that, because low-income renters are highly mobile within cities, gentrification affects incumbent renters primarily by changing the characteristics of other neighborhoods in their choice sets. Our results imply that where low-income renters lived within US metro areas mattered comparatively less than which US metro areas they lived in. Policies favoring neighborhood incumbency may be poorly aligned with how gentrification affects the large majority of low-income renters.

*Any views expressed are those of the authors and not those of the U.S. Census Bureau. The Census Bureau has reviewed this data product to ensure appropriate access, use, and disclosure avoidance protection of the confidential source data used to produce this product. This research was performed at a Federal Statistical Research Data Center under FSRDC Project Number 2358 (CBDRB-FY24-P2358-R10957, CBDRB-FY24-P2358-R10936). This research uses data from the Census Bureau's Longitudinal Employer Household Dynamics Program, which was partially supported by National Science Foundation grants SES-9978093, SES-0339191 and ITR-0427889; National Institute on Aging Grant AG018854; and grants from the Alfred P. Sloan Foundation. This research was supported by The Horowitz Foundation for Social Policy and a Stone Research Grant from Harvard Kennedy School's James M. and Cathleen D. Stone Program in Wealth Distribution, Inequality, and Social Policy. Additional acknowledgments to come.

1. Introduction

Beginning in the 1990s and intensifying after the year 2000, gentrification transformed the socioeconomic composition of many American inner city neighborhoods (Baum-Snow and Hartley 2020; Couture and Handbury 2023). Between 2000 and 2017, the share of residents with a college degree in census tracts near their metro areas' central business districts rose by an average of 15 percentage points from a baseline of 24 percent, more than twice the 7-percentage-point increase in suburban neighborhoods. There are competing interpretations of this transformation's consequences for incumbent residents. Studies finding limited mobility responses to gentrification suggest it may have had limited effects on incumbent residents (Freeman and Braconi 2004; McKinnish, Walsh, and White 2010; Ellen and O'Regan 2011; Brummet and Reed 2021), while research on declining urban affordability and changing urban amenities points to potentially large welfare effects for incumbent residents (Diamond 2016; Almagro and Domínguez-Iino 2022; Baum-Snow and Duranton 2025; Couture et al. 2023).

In this paper, we study how gentrification affected incumbent low-income working renters ages 25 to 65 who, in 2010, lived in neighborhoods with initially low shares of college graduates. While we consider alternative definitions, we measure gentrification primarily as increases in a neighborhood's share of college-educated adults. We pair a comprehensive descriptive analysis with a dynamic structural model of residential and workplace choice to assess how renters in gentrifying neighborhoods fared relative to comparable renters in other initially low-college-share neighborhoods in the same metro that did not gentrify. Our central finding is that, because low-income renters are highly mobile within cities, their welfare depended less on whether their initial neighborhood gentrified than on how affordability, amenities, and employment opportunities evolved throughout their metro-area. This distinction between neighborhood-level exposure to gentrification and changes in metro-area choice sets helps reconcile limited mobility responses with the research pointing to potentially large welfare effects from gentrification.

Recent increases in neighborhoods' college-educated shares have been associated with changes in local amenities (Hoelzlein 2023; Su 2022; Ellen, Horn, and Reed 2019), housing costs (Baum-Snow and Duranton 2025; Guerrieri, Hartley, and Hurst 2013), and employment opportunities (Lester and Hartley 2014; Meltzer and Ghorbani 2017; Brummet and Reed 2021). Incumbent renters of these neighborhoods may have benefited if they valued the associated amenity and employment changes more than the accompanying rent increases. But if neighborhoods changed in ways they did not value, gentrification may have lowered their welfare if moving is costly. However, as emphasized by Vigdor (2002), if moving costs are low, renters could have reoptimized by relocating to neighborhoods better aligned with their preferences. In that case, their welfare depended less on whether their initial neighborhood gentrified than on whether the broader set of accessible neighborhoods remained affordable and desirable. We find that this latter case is more empirically relevant.

We base our analysis on newly available administrative U.S. Census data that track the earnings, workplaces, and residential addresses of more than 3 million low-income urban renter households across large U.S. metro areas. We carefully link person-level records from the Census Bureau's Master Address File (MAF), the American Community Survey (ACS), the Longitudinal Employer-Household Dynamics (LEHD) database, and real estate data from CoreLogic, among other administrative sources. We first use our panel to document how low-income renters' neighborhood characteristics and residential choices varied with gentrification. Then, because our descriptive findings admit competing welfare interpretations, we estimate a dynamic model of neighborhood and workplace choice to quantify the welfare effects of gentrification for incumbent low-income renters. Throughout the analysis, we report results separately for Black and non-Black households, motivated by longstanding concerns that gentrification has disproportionately harmed Black residents (Freeman 2006; Pattillo 2010).

We begin by showing that our sample of low-income renters is highly mobile. The baseline probability of moving census tracts between 2010 and 2011 was 18.49% for non-Black and 18.69% for Black household heads, nearly twice the 10.25% rate for all household heads regardless of income, race, or homeownership status. These move probabilities increase to 69.84% for non-Black and 72.98% for Black low-income renter household heads over 2010-2019, compared with 52.07% for all household heads. Notably, these high mobility rates are largely insensitive to whether one's origin neighborhood gentrified. Low-income Black renters in the top decile of gentrifying tracts were only 1.16 percentage points more likely to move tracts between 2010 and 2019 than the average low-income Black renter household head. For non-Black renters, this difference is just 0.19 percentage points.

The small mobility responses are consistent with a large literature examining residential mobility and gentrification, including Brummet and Reed (2021), who link households across the 2000 Decennial Census and the 2010–2014 ACS.[1] We build on their analysis in two ways. First, because their gentrification measures rely on noisy five-year ACS tract estimates, classical measurement error may attenuate their descriptive estimates, leaving open the possibility of large mobility responses to gentrification. We instead link the MAF to the Census Bureau's demographic surveys to construct precise measures of tract-level sociodemographic change and use these measures to quantify the attenuation bias in descriptive regressions based on ACS aggregates. Second, we test whether the small average mobility responses mask meaningful heterogeneity across neighborhood and renter characteristics. However, even with our more precise measures, we find small average mobility responses across a wide range of neighborhood and renter characteristics, including settings where larger responses might have been expected, such as tracts with inelastic housing supply (Baum-Snow and Han 2023), tracts not covered by rent control ordinances, and among elderly renters and those with long prior residential tenure.

We next examine whether gentrification was associated with changes in incumbent renters' economic outcomes. Using linked employer-employee records from the LEHD, we study both proximity to employment opportunities and realized employment, commuting, and earnings outcomes. We find small changes in proximity to employment opportunities, with modest improvements for non-Black low-income households and negligible changes for Black low-income households. We do, however, find a meaningful association with realized outcomes among Black households. In particular, Black low-income renters whose initial tract's college share doubled between 2010 and 2019 were 10 percent, or one percentage point, less likely to have positive earnings recorded in the LEHD in 2019 than comparable Black renters in non-gentrifying

[1] Additional research focusing on gentrification's impact on the propensity of incumbent residents to leave their home neighborhoods not yet cited includes Ding, Hwang, and Divringi (2016), Dragan, Ellen, and Glied (2020), Baum-Snow, Hartley, and Lee (2019), Beauregard (2024), and Pennington (2021). Baum-Snow, Hartley, and Lee (2019) examine the impact of neighborhood change on children's long-run outcomes while Beauregard (2024) uses rich administrative panel data in Canada and finds that the effects of gentrification on residential mobility vary across Canadian cities.

tracts. Our findings are generally consistent with Meltzer and Ghorbani (2017), who find little evidence of higher employment rates in gentrifying neighborhoods in the New York metro area, and with Brummet and Reed (2021), who find little evidence that gentrification altered incumbent adults' employment or incomes. We examine both proximity to employment opportunities and realized economic outcomes for the same samples, but do find meaningful associations with realized outcomes among Black low-income renters.

We end the descriptive analysis by characterizing the neighborhoods incumbent renters occupied in 2019, separately comparing those who remained in their origin tract since 2010 with those who moved. Stayers in gentrifying tracts experienced much larger rent increases than movers. Relative to renters from non-gentrifying tracts, a doubling of an incumbent's initial tract college-graduate share is associated with roughly 25% higher rent growth for stayers, but only 3% higher rent growth for movers. More generally, movers from gentrifying tracts tended to reach destination tracts that were observably similar to those reached by movers from non-gentrifying tracts. Destination tracts had comparable median rents, college-graduate shares, and, for non-Black households, white population shares. One exception is that Black low-income renters leaving gentrifying tracts moved to neighborhoods with significantly lower white population shares than Black movers from non-gentrifying tracts, consistent with racial mobility networks shaping migration patterns in gentrifying neighborhoods (Ferreira, Kenney, and Smith 2023). Movers from gentrifying tracts also reached destinations with marginally employment opportunities.

Our descriptive findings admit two broad interpretations of the welfare effects of gentrification for incumbent low-income renters. If moving costs are low, observed high mobility and similar destination tract choices suggest that renters in gentrifying and non-gentrifying tracts could readily access similar neighborhoods, limiting the welfare differences between these groups. Alternatively, observed high residential mobility may reflect frequent renter- and neighborhood-specific shocks rather than directed responses to neighborhood change (Collinson et al. 2024; Desmond 2017; DeLuca, Wood, and Rosenblatt 2019). In this case, moving costs may still be sizeable, helping explain why incumbent renters responded only weakly to gentrification. Under this second interpretation, the welfare effects for renters who stayed in their origin neighborhoods are ambiguous as they depend on their willingness to pay for the changes in neighborhood characteristics that accompanied gentrification throughout 2010-2019.

To quantify the welfare effects of gentrification and distinguish between these competing interpretations for incumbent renters, we estimate a dynamic discrete neighborhood choice model tailored to low-income working renters (Bayer et al. 2016; Davis et al. 2021; Almagro and Domínguez-Iino 2022; Davis, Gregory, and Hartley 2023). In each period, forward-looking renters choose where to live and where to work to maximize expected present-discounted utility, subject to rich moving costs. Flow utility combines expected housing and non-housing consumption, neighborhood amenities, and neighborhood-specific capital that accumulates with residential tenure, capturing the value of local familiarity and social ties (Kan 2007; Koenen and Johnston 2024). To capture differences in neighborhood employment opportunities and commuting costs, we embed a repeated static workplace choice problem within renters' dynamic neighborhood choice problem. This feature extends existing dynamic neighborhood choice frameworks by incorporating microfounded, time-varying neighborhood-level measures of employment access into households' flow utilities (Tsivanidis 2022).[2]

We further extend the model by allowing rents paid by incumbents to vary with the length of residential tenure in places with binding rent control ordinances, capturing the partial insulation from market rent growth that incumbent renters enjoy (Diamond, McQuade, and Qian 2019). Without allowing for tenure-dependent rents, we could mistake continued residence in neighborhoods with rising market rents for strong neighborhood attachment and understate renters' disutility from higher rents. We also let moving costs vary with both physical distance and differences in college-graduate shares across tracts. We refer to the latter as "social moving costs". These costs capture the idea that, while low-income renters may value living in higher-socioeconomic-status neighborhoods, they may face non-financial frictions that make such moves unattractive (Bergman et al. 2023).

Our welfare calculations depend on the mean utility renters obtain from living in each neighborhood. These mean utilities determine whether a gentrifying neighborhood raised or lowered welfare for a low-income renter constrained to stay there. To understand why gentrification raised or lowered welfare, we separately identify how renters value the components of mean utility, including tract rents, employment access, amenities causally tied to neighborhoods' college-graduate shares, and potentially correlated unobserved amenities not causally tied to these shares (Lee and Lin 2018). We estimate renters' preferences by comparing changes in neighborhood choices across cohorts of low-income renters and instrumenting for changes in observed neighborhood characteristics over 2010-2019. We use an Euler equations in conditional choice probabilities estimator from the dynamic discrete choice literature (Scott 2013; Kalouptsidi, Scott, and Souza-Rodrigues 2021).[3] This approach yields estimating equations that are linear in the model parameters, which we estimate by two-stage least squares.

We instrument for changes in job market access using national growth in tradable six-digit NAICS industries, weighted by each tract's initial industry employment shares. Building on Severen (2023), Baum-Snow, Hartley, and Lee (2019), and Baum-Snow and Han (2023), our instrument uses sample-specific six-digit industry variation rather than two-digit sector employment, helping separate tradable employment shocks from changes in local consumer demand (Couture and Handbury 2020) while providing enough cross-industry variation for us to treat initial employment shares as potentially correlated with evolving exogenous amenities (Borusyak, Hull, and Jaravel 2022). To instrument for changes in a neighborhood's college-graduate share, we interact distance-weighted exposure to nearby neighborhoods' college-graduate shares with changes in metro-wide demand among college graduates, following evidence that low-income neighborhoods near high-income neighborhoods appreciate most during positive labor-demand shocks (Guerrieri, Hartley, and Hurst 2013). Finally, analogously to Diamond (2016), we interact these predicted changes in college-shares with tract-level measures of developed land to generate variation in rents, given the tight relationship between the extent of developed land and local housing supply elasticities

[2]These neighborhood-level measures of employment access are the same measures we use in our descriptive analysis. In ongoing and contemporaneous work, Warnes (2024) similarly incorporates a repeated-static workplace choice problem into a dynamic neighborhood choice model to evaluate the welfare impact of transportation infrastructure improvements in Buenos Aires.

[3]ECCP estimation has been used in a variety of applied settings, from choices over agricultural land use (Scott 2013) and new technology adoption (Groote and Verboven 2019) to 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-Iino 2022).

(Baum-Snow and Han 2023).

We use the estimated model to quantify the welfare impacts of gentrification from the perspective of incumbent low-income renters in 2010. For each tract, we compare expected welfare under the observed evolution of tract characteristics with expected welfare under a counterfactual steady-state that holds tract characteristics fixed at their 2010 levels. These comparisons reveal only small differences between initially low-college-share tracts that did and did not gentrify, suggesting limited within-city differences in the welfare impacts of gentrification for incumbent low-income renters. Specifically, we find that low-income renters initially living in tracts that saw their college-graduate-share double between 2010 and 2019 had only x% lower expected welfare than low-income renters initially living in comparable tracts in the same metro area whose college-graduate-share remained stable.[4] These welfare losses are slightly larger for Black renters, as they valued the changes in origin-neighborhood characteristics that accompanied gentrification less than non-Black renters did.

Lastly, we decompose tract-level differences in expected welfare into changes in low-income renters' option value of moving and changes in expected welfare from staying in their origin tract. We define the option value of moving as roughly the average welfare difference from moving to another tract in the renter's choice set relative to not moving. It captures how much access to the broader neighborhood choice set insulates renters from adverse changes in their origin tract, and is thus higher when nearby alternative neighborhoods are attractive and lower when moving costs are high. We find that declines in the present discounted value of staying in a gentrifying tract are typically offset by higher option values from moving, suggesting that access to alternative neighborhoods constrains the negative welfare effects of a renter's origin tract gentrifying. Even where this compensating effect is weaker in gentrifying tracts, option values also decline in non-gentrifying tracts, leaving within-metro welfare differences modest. These patterns suggest that the welfare consequences of gentrification operate more through differences in metro area neighborhood choice sets than through differences across origin tracts within metros. Our results suggest that where low-income renters lived within US metro areas mattered comparatively less than which US metro area they lived in.

Our findings connect to a growing literature on the dynamic nature of neighborhoods, in which a neighborhood's trajectory often says little about what happens to the people who lived there. Garin et al. (2025) document this for concentrated poverty, finding that residents of poor tracts are highly mobile and tend to move away as their earnings grow while the tracts they leave stay poor. In ongoing and closely relate work, Baum-Snow et al. (2025) examine how Texan renters arbitrage across neighborhoods to avoid rent increases, while Baum-Snow, Hartley, and Lee (2019) study how neighborhood change affects incumbent children in tracts experiencing high-skill labor-market access shocks. We take a distinctly retrospective approach to studying gentrification, studying its effects on incumbent renters without focusing on a particular source of neighborhood change.

Policies favoring neighborhood incumbency may be poorly aligned with how gentrification affects the large majority of low-income renters. Rent control, for example, may improve welfare for incumbent renters who remain but reduce welfare for prospective renters and movers seeking access to affordable neighborhoods (Diamond, McQuade, and Qian 2019). Because our findings suggests that renters' welfare

[4] We are currently finalizing our welfare estimates and will update these paragraphs soon.

depends heavily on the broader set of neighborhoods they can access within a metro area, policies that improve metro-wide affordability may better target the margins that shape most renters' welfare.

2. Data

Summary coming soon. See appendix for complete details.

3. Descriptive Statistics

3.1. Summary Statistics

TABLE 1. Sample Summary Statistics

A. Household Head Characteristics, 2010

	Low-Income Renters	
	Black	Non-Black
Adj. HH Income	25,050	26,590
	(18,350)	(17,920)
College Degree	0.0698	0.0681
	(0.255)	(0.252)
Age	41.14	41.78
	(10.50)	(10.59)
Parent	0.4406	0.4719
	(0.4965)	(0.4992)
Female	0.6300	0.5183
	(0.4828)	(0.4997)
White	—	0.8555
	(—)	(0.3516)
Immigrant	0.1335	0.4475
	(0.3401)	(0.4972)
Hispanic	0.09036	0.5948
	(0.2867)	(0.4909)
Tenure Length (years)	4.975	5.200
	(3.355)	(3.371)
Commute Distance (miles)	9.61	9.78
	(9.19)	(9.55)
Commute Time (minutes)	30.1	25.1
	(24.2)	(21.1)

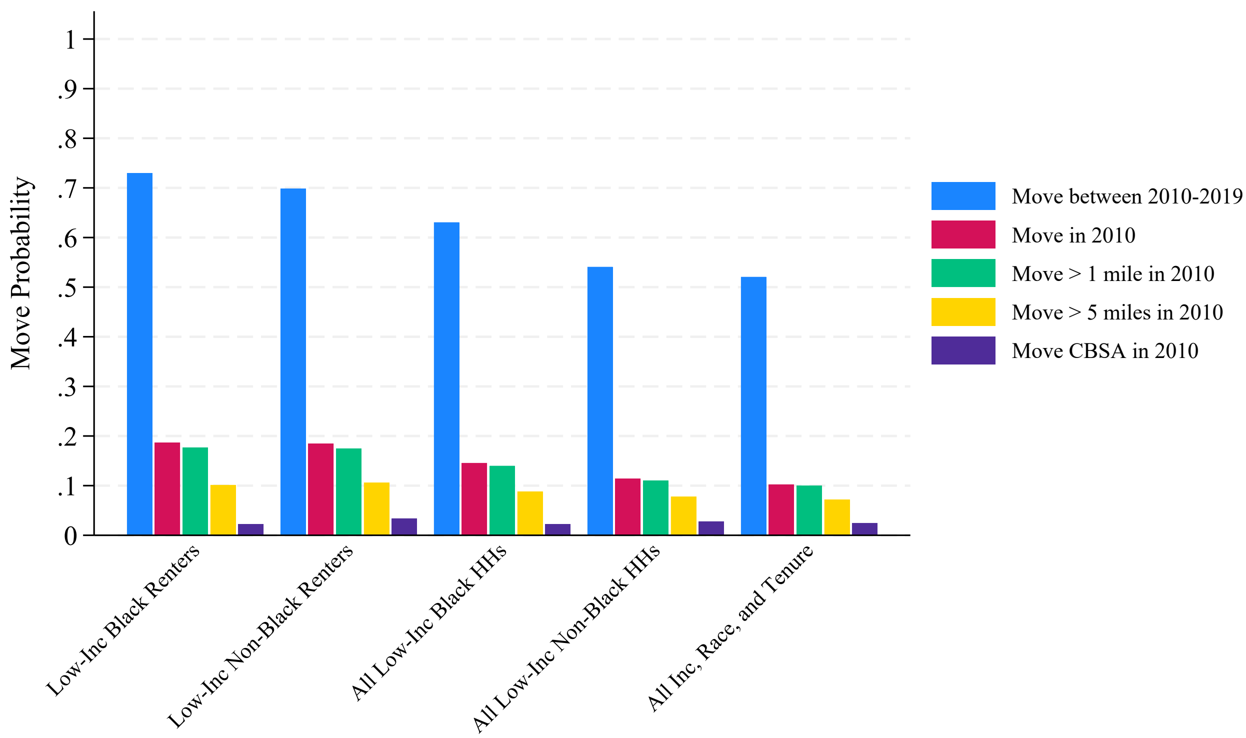
B. Household Tract Characteristics, 2010

	Low-Income Renters	
	Black	Non-Black
CBD Distance (miles)	8.406	8.481
	(5.208)	(5.260)
CBD Distance (pop. density)	0.2010	0.2083
	(0.1381)	(0.1404)
Hedonically Adjusted Rents	502.5	554.2
	(141.7)	(147.1)
College Share	0.0954	0.0947
	(0.0328)	(0.0354)
White Share	0.2651	0.5520
	(0.2317)	(0.2327)
Black Share	0.5315	0.1439
	(0.3196)	(0.1869)
Hispanic Share	0.2631	0.5153
	(0.2808)	(0.3032)
Immigrant Share	0.2752	0.4368
	(0.2430)	(0.2456)
Total Adult Population	2,129	2,379
	(912.7)	(922.8)

Number Black Households: 351,000
Number Non-Black Households: 517,000

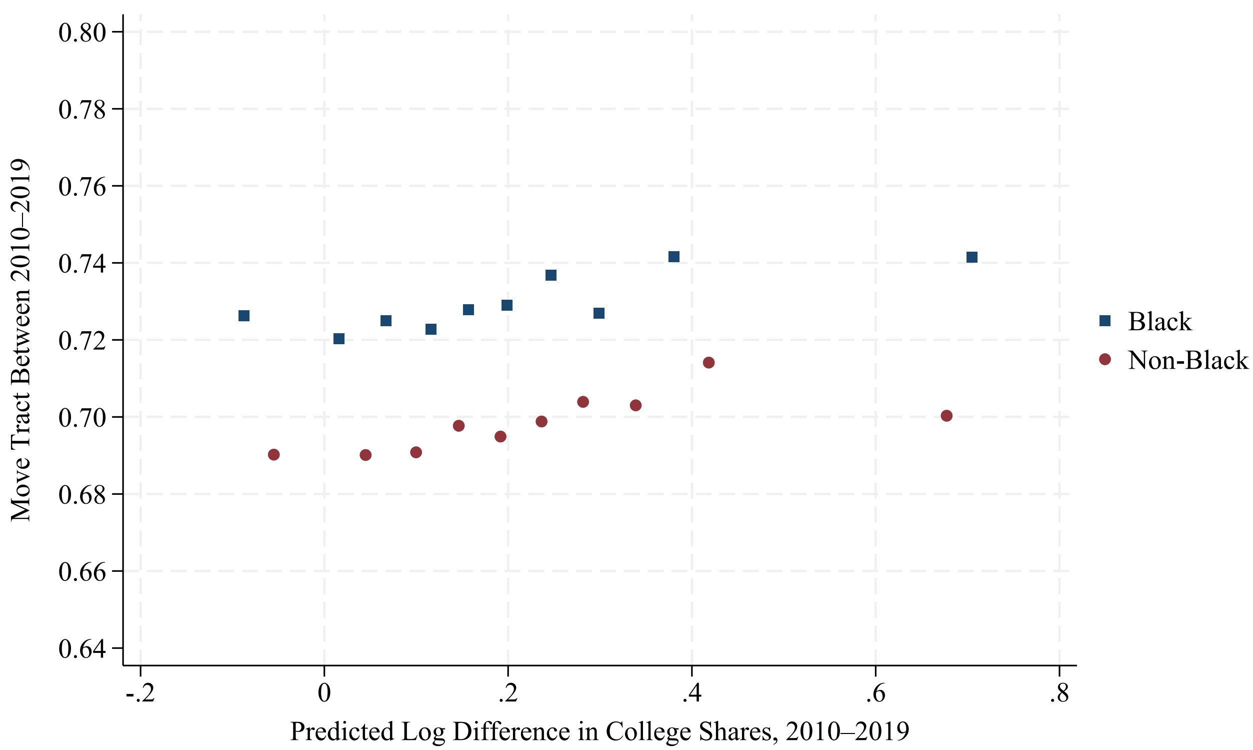
Notes: Summary statistics are for incumbent low-income urban renters living in tracts among the bottom quartile of their CBSA's college-attainment distribution in 2010 (the "Baseline mobility sample" in Table A1). Panel A reports household head characteristics while Panel B reports characteristics of the household's 2010 census tract. Entries are means, with standard deviations in parentheses. Dollar amounts are in 2010 USD. Parental status, commute time, and commute distance is computed only for the subsample that we can link to the ACS. CBD distance is measured in miles to the Metropolitan Division's CBD and, separately, by the population share residing closer to the Metropolitan Division's CBD than the tract in 2010. We detail hedonically adjusted rents in Appendix A.3 and adjusted household income in Footnote 41. Tract sociodemographic shares and total adult population are constructed using the de-biased tract aggregates detailed in Appendix A.3. All values are rounded according to Census guidelines, which depend on underlying sample sizes and variable type.

Figure 1. Baseline Residential Mobility



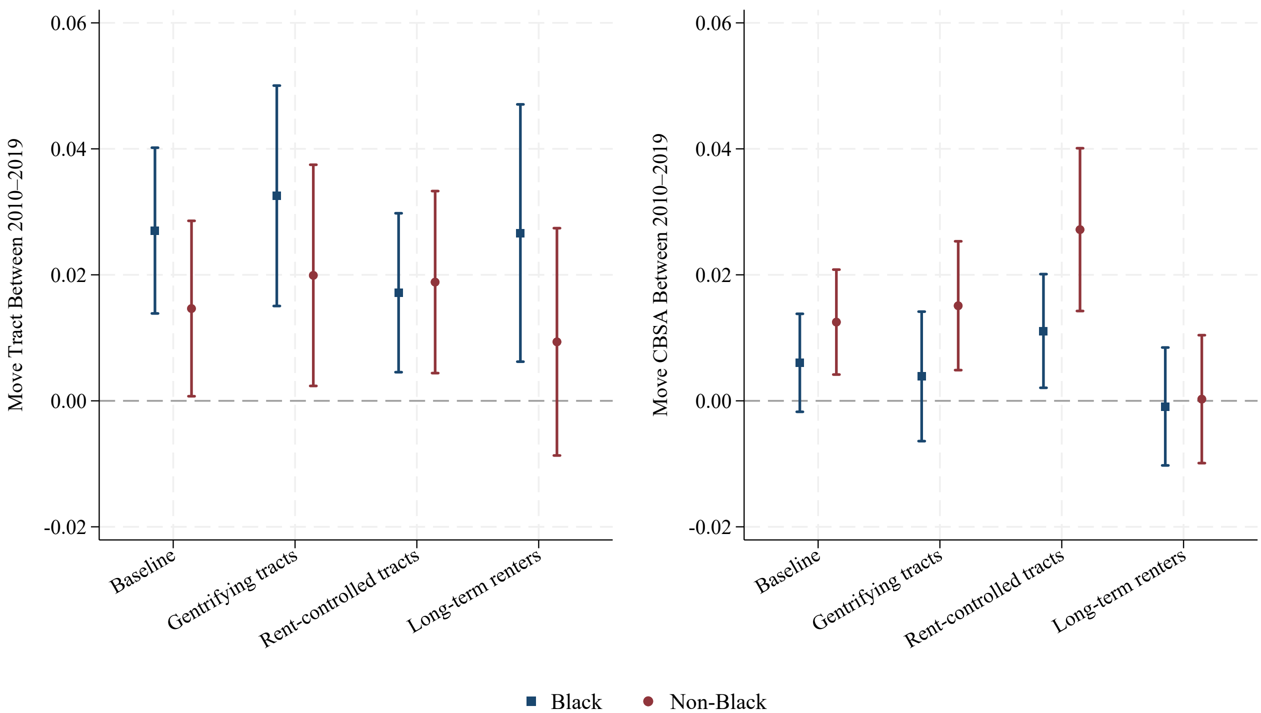
Notes: Figure 1 summarizes residential mobility for households observed in 2010. The leftmost bars report the share of household heads who lived in a different census tract in 2019 than in 2010; the remaining bars report shares who lived in a different census tracts in 2011 than in 2010. "Low-Inc Black (Non-Black) Renters" are low-income urban renters living in tracts in the bottom quartile of their CBSA's college-attainment distribution in 2010. "All Low-Inc Black (Non-Black) HHs" are all low-income household heads in 2010, regardless of homeownership status. "All Inc, Race, and Tenure" includes all 2010 household heads, irrespective of income, race, or homeownership status. These samples correspond to those defined after applying restrictions (h.i), (f), and (d) in Table A1, respectively. We compute residential mobility statistics only when the household head's census tract is observed in both years relevant to a given bar (2010 and 2019 for the leftmost bars; 2010 and 2011 for the others). We report sample sizes and additional baseline mobility statistics in Table A13.

Figure 2. Gentrification and Residential Mobility



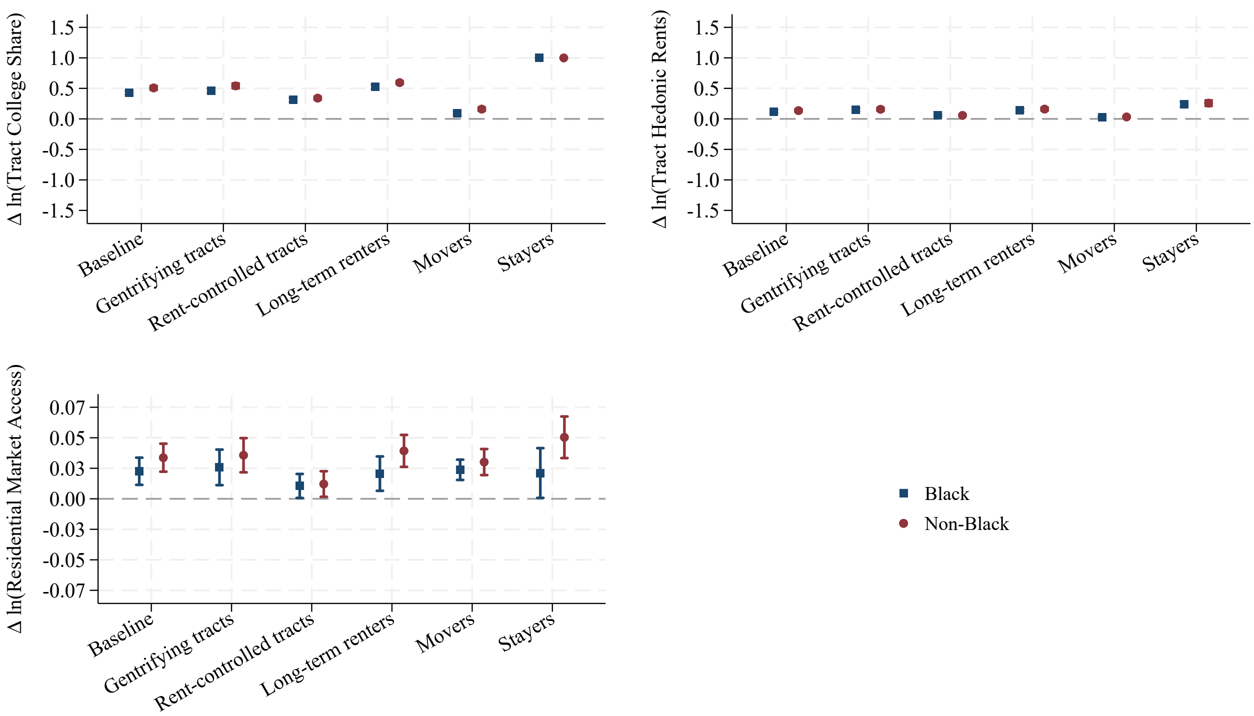
Notes: Figure 2 shows the probability incumbent renters move census tracts between 2010-2019 as a function of their origin tract's predicted log change in the share of adults with at least a bachelor's degree. The points plot conditional means estimated from a regression that includes CBSA fixed effects and the full set of baseline controls listed in Table **??** (Cattaneo et al. 2024). We describe the predicted log change in college shares in the main text. The sample comprises urban low-income renter households observed in both 2010 and 2019, defined after applying restriction (h.ii) in Table A1.

Figure 3. Gentrification and Residential Mobility Coefficient Estimates



Notes: Figure 3 plots coefficient estimates for our gentrification measure from regression equation **??**. In the left plot, the dependent variable is an indicator equal to one if the incumbent household head moved to a different census tract by 2019, whereas in the right plot it equals one if the household head moved to a different CBSA by 2019. We instrument for gentrification using changes in the adult college-educated population, as described in the main text. All specifications include CBSA fixed effects and the full set of baseline controls listed in Table **??**. We define gentrifying tracts as those whose college share rose from 2010 to 2019; rent-controlled tracts as those covered by a rent-control ordinance in 2010; and long-term renters as households residing in the same tract since at least 2005. We display 95% confidence intervals based on robust standard errors clustered at the origin census-tract level (Abadie et al. 2023). The sample comprises urban low-income renter households observed in both 2010 and 2019, defined after applying restriction (h.ii) in Table A1.

Figure 4. Gentrification and Experienced Neighborhood Characteristics Coefficient Estimates



Notes: Figure 4 plots coefficient estimates for our gentrification measure from regression equation **??**. In the top-left plot, the dependent variable is the log difference in the share of adults with a college degree between each household's 2019 and 2010 census tracts. In the top-right plot, the dependent variable is the log difference in hedonic rents between each household's 2019 and 2010 census tracts. In the bottom-left plot, the dependent variable is the log difference in residential market access between each household's 2019 and 2010 census tracts. Movers are household heads observed in a different census tract in 2019 than in 2010. Stayers are household heads observed in the same census tract in 2019 as in 2010. Otherwise, the plots are analogous to those in Figure 3.

4. A Dynamic Model of Neighborhood and Workplace Choice

Among papers in the dynamic discrete choice literature, our setup is most similar to the neighborhood demand model of Almagro and Domínguez-Iino (2022), which analyzes the endogenous formation of horizontally differentiated private consumption amenities in Amsterdam's 2010–2019 tourism boom. Full preamble coming soon.

4.1. Households and Timing of Choices

Low-income renter households are indexed by i and differ ex-ante by race, $k \in \{\text{Black}, \text{Non-Black}\}$. Access to employment opportunities and preferences over neighborhood characteristics can vary by household race. Households also differ by their current residential neighborhood, $n_{i,t-1}$, and the length of residential tenure in this neighborhood, $\tau_{i,t-1}$. These controlled state variables, $x_{i,t} \equiv (n_{i,t-1}, \tau_{i,t-1})$, influence households' choices by affecting the utility cost of moving from their current neighborhood.[5] We consider households choosing among neighborhoods in their current city, c, and an outside option of leaving their city, OO^c. We denote these city-specific choice sets as $\mathcal{N}^c \equiv \{OO^c, 1^c, \ldots, N^c\}$ and abstract from neighborhood choices between cities.

Time is discrete and indexed by t. In each period, households receive idiosyncratic preference shocks ε_{int} for residential neighborhoods within their city and for their outside option of leaving their city. They also receive idiosyncratic productivity shocks $z_{im\iota t}$ across workplace locations, $m \in \mathcal{N}^c$, and industries, $\iota \in \mathrm{I}$, where I denotes the set of industries. After first observing their residential preference shocks, households choose their residential neighborhood n_{it} and then decide on an optimal level of housing consumption H^k_{nt}. Following this, they observe their idiosyncratic workplace productivity shocks and choose their workplace location and industry.[6] The timing of these choices is summarized in Figure 5.

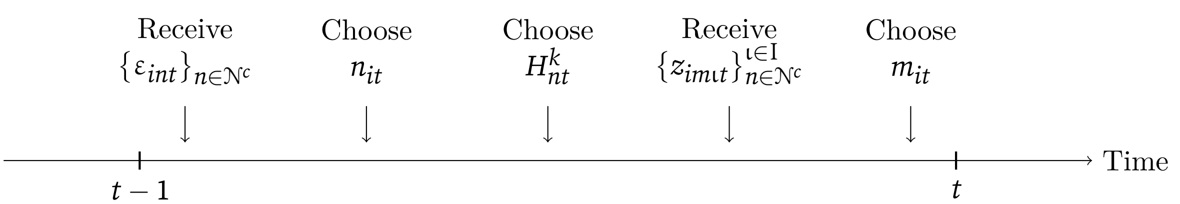


FIGURE 5. Timeline of Household Choices t

In the remainder of this section, we detail the periodic household problem working backwards from their workplace choice.

[5] The individual-level state variables, x_{it}, are "controlled" in that their evolution is fully determined by the household's choices.

[6] We separate households' residential location and workplace choices to ease exposition. Our framework is isomorphic to a framework where households jointly choose residential and workplace locations as in Ahlfeldt et al. (2015).

4.2. Workplace Choice

Upon making their residential neighborhood and housing consumption choices, households receive idiosyncratic productivity shocks across workplace locations and industries, $z_{im\iota t}$. Conditional on living in neighborhood n, households observe these productivity shocks and choose their workplace location and industry to maximize their commute time–discounted income:

$$I_{int} \equiv \max_{m,\iota} \frac{z_{im\iota t}}{d^k_{nmt}} \, w^k_{m\iota t}$$

where $w^k_{m\iota t}$ is the wage offered in workplace tract m, industry ι, and period t to type-k households measured in efficiency units. $d^k_{nmt} > 0$ is the cost for a type-k household to commute between neighborhoods n and m in period t also measured in efficiency units.[7] We assume that households spend a fixed amount of time each day working or commuting, so d^k_{nmt} effectively discounts the total wage offered to household i in workplace m and industry ι: $z_{im\iota t} \cdot w^k_{m\iota t}$. We assume that $z_{im\iota t}$ is drawn independently from a Frechet distribution with shape parameter ϵ^{kc}_t. These shape parameters are specific to each city, year, and racial group, which we make explicit with the superscripts k, c and subscript t.[8] The expected income for a type-k household living in tract n at time t is therefore

$$\bar{I}^k_{nt} = \Gamma\left(1 - \frac{1}{\epsilon^{kc}_t}\right) \cdot \left(\mathrm{RMA}^k_{nt}\right)^{1/\epsilon^{kc}_t}$$

where $\Gamma(\cdot)$ is the gamma function and $\mathrm{RMA}^k_{nt} \equiv \sum_{\iota \in \mathrm{I}} \sum_{m \in \mathcal{N}^c} \left(\frac{w^k_{m\iota t}}{d^k_{nmt}}\right)^{\epsilon^{kc}_t}$ is a summary measure of access to employment, which we follow the literature in terming residential market access. We derive these expressions and describe how we construct their empirical analogs in Appendix B.1.

4.3. Neighborhood Choice

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

$$\max_{\{n \in \mathcal{N}^c\}^\infty_t} \quad \mathbb{E}\left[\sum_{t'=t}^{\infty} \delta^{t'-t} \cdot u^k_n\left(s^c_{it'}\right) \,\middle|\, \mathcal{I}_{it}\right] \tag{1}$$

where δ is a known discount factor and s^c_{it} is a vector of state variables that determine household i's flow utility u^k_n from choosing neighborhood n. s^c_{it} includes the measures of expected income, $\bar{I}^k_{nt}$, derived in Section 4.2. $\mathbb{E}\left[\cdot|\mathcal{I}_{it}\right]$ denotes the expectation operator conditioned on household i's information set at time t. 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

[7] A note here about why we allow the cost of commutes between tracts to differ by race (report summary statistics on differences in commute times across races within tract pairs.).

[8] Assuming the shocks are independent across years implies negligible costs to switching jobs. I should highlight the high job switching rates reported in Appendix ZZ.

current and past state variables that households may use to form expectations over their future evolution. We specify households' belief formation in Section 5.1.

State Variables. Households' flow utilities depend on the vector of state variables $s^c_{it} \equiv \left(x_{it}, \varepsilon_{it}, \omega^{kc}_t, \xi^{kc}_t\right)$, where $(x_{it}, \varepsilon_{it})$ are household-level observable and unobservable state variables, respectively. By contrast, $\left(\omega^{kc}_t, \xi^{kc}_t\right)$ are city-specific observable and unobservable state variables, respectively. The household-level observable state variables, x_{it}, include the length of households' residential tenure and their neighborhood choice in the previous period, $x_{it} = (n_{it-1}, \tau_{it-1})$. We assume the household's unobservable state, $\varepsilon_{it} \equiv \{\varepsilon_{int}\}_{n \in \mathcal{N}^c}$, is i.i.d. across households, neighborhoods, and time. We conceptualize ε_{it} as a vector of unobserved-to-the-econometrician time-varying household and neighborhood-specific preference shocks. As is common, we assume that ε_{it} is distributed according to a type I extreme value distribution with scale parameter σ^k.

ω^{kc}_t denotes observable city- and type-specific state variables. It includes vectors of housing costs that can vary by the length of residents' tenure in a neighborhood, $r_{\tau nt}$, shares of college graduates across neighborhoods, $\frac{\text{Coll}_{nt}}{\text{Pop}_{nt}}$, each neighborhoods' commute time–discounted expected incomes, $\bar{I}^k_{nt}$, and an index for each period to explicitly incorporate nonstationarity:

$$\omega^{kc}_t = \left(\{r_{\tau nt}\}, \left\{\frac{\text{Coll}_{nt}}{\text{Pop}_{nt}}\right\}, \left\{\bar{I}^k_{nt}\right\}, t\right).$$

Finally, ξ^{kc}_t is a city-specific vector of unobservable time-varying neighborhood-level amenity valuations among type k households. For example, ξ^{kc}_t could include time-varying valuations among type-k residents for suburban life, independent of gentrification. To facilitate exposition, we define $\bar{\omega}^{kc}_t \equiv \left(\omega^{kc}_t, \xi^{kc}_t\right)$ as the vector containing both observable and unobservable city-specific state variables.

Flow Utility. Flow utility net of moving costs for a type-k household with residential tenure length τ that chooses neighborhood n can be represented as

$$A^k_{nt} \cdot Q^k_{\tau nt} \cdot \tau^{\beta^k_\tau} \cdot \exp(\varepsilon_{int}), \tag{2}$$

where A^k_{nt} is a type-specific valuation of amenities in neighborhood n and $Q^k_{\tau nt}$ is a consumption composite that is Cobb–Douglas over nonhousing consumption, $C^k_{\tau nt}$, and housing consumption, $H^k_{\tau nt}$:

$$Q^k_{\tau nt} \equiv \left(C^k_{\tau nt}\right)^{\beta^k_C} \left(H^k_{\tau nt}\right)^{1-\beta^k_C}. \tag{3}$$

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

$$C^k_{\tau nt} \geq \bar{I}^k_{nt} - r_{\tau nt} \cdot H^k_{\tau nt},$$

where time-varying neighborhood-level rents can depend on the length of residents' tenure as,

$$r_{\tau nt} = r_{nt} \cdot h_n\left(\tau, \{\bar{\omega}_t^{kc}\}_{t-\tau}^{t}\right). \tag{4}$$

Equation 4 decomposes rents charged to residents in neighborhood n in period t into current market-level rents, r_{nt}, and a component that depends on the length of residents' tenure and a vector of state variables in neighborhood n: $h_n\left(\tau, \{\bar{\omega}_t^{kc}\}_{t-\tau}^{t}\right)$.[9] $\{\bar{\omega}_t^{kc}\}_{t-\tau}^{t}$ can contain information on the prevalence of any local rent control ordinances, changes in historical market rents, and other non-market forces influencing how charged rents vary with the length of neighborhood tenure for low-income renters. Incorporating tenure-dependent rents helps us distinguish between households' disutility from paying higher rents and utility from remaining in their home neighborhood all else equal, $\tau^{\beta_\tau^k}$. We detail the empirical construction of $r_{\tau nt}$ in Appendix A.3.

Type-specific neighborhood amenities are

$$A_{nt}^k \equiv \left(\frac{\text{Coll}_{nt}}{\text{Pop}_{nt}}\right)^{\beta_A^k} \exp\left(\xi_{nt}^k\right).$$

We decompose unobserved neighborhood- and period-specific amenities, ξ_{nt}^k, into time-invariant neighborhood-specific components, time-varying but neighborhood-invariant components, and neighborhood-specific time-varying components:

$$\xi_{nt}^k \equiv \alpha_n^k + \alpha_t^k + \bar{\xi}_{nt}^k.$$

Appendix B.2 shows that by taking logs of equation 2, incorporating moving costs, $MC_t^k(n_t, n_{it-1})$ (defined below), solving for optimal housing consumption, and substituting in the amenity specifications yields the following expected flow utility specification for a type-k household choosing neighborhood n with state s_{it}^c:

$$\begin{aligned} u_n^k\left(s_{it}^c\right) &= \tilde{\alpha}_n^k + \tilde{\alpha}_t^c + \tilde{\beta}_I^k \ln\left(\bar{I}_{nt}^k\right) + \tilde{\beta}_r^k \log\left(r_{\tau nt}\right) + \tilde{\beta}_A^k \ln\left(\frac{\text{Coll}_{nt}}{\text{Pop}_{nt}}\right) \\ &+ \tilde{\beta}_\tau^k \ln(\tau_{it}) + MC_t^k(n_t, n_{it-1}) + \tilde{\xi}_{nt}^k + \varepsilon_{int} \end{aligned} \tag{5}$$

Flow utility is expected in that it is the value households expect to obtain before they realize their period-specific workplace-tract productivity shocks. It is this flow utility specification that is relevant for households' neighborhood choice and thus for our structural estimation of households' preferences over neighborhood characteristics.

Moving Costs. If a type-k household decides to leave its current neighborhood for another neighborhood in the same city, it incurs a nonmonetary moving cost, $MC_t^k(n_{it}, n_{it-1})$. This nonmonetary moving cost comprises a city-specific fixed disutility from moving, the physical straight-line distance between the

[9]We constrain $h_{OO^c}\left(\tau, \{\bar{\omega}_t^{kc}\}_{t-\tau}^{t}\right) = 1$. This constraint assumes households residing outside their associated urban core always pay market rents, regardless of their neighborhood tenure. This assumption is motivated by the small share of rural residential units located in tracts with binding rent control regulations or government subsidies Check.

household's origin and destination neighborhoods, and the *social distance* (defined below) between these two neighborhoods. Conversely, if a type-k household decides to leave its city entirely, it incurs a single, city-specific fixed cost. Specifically,

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

where MC^{kc} and MC_{OO}^{kc} are the city-specific fixed intensive- and extensive-margin moving costs. $\boldsymbol{d}(n_t, n_{t-1})$ is a vector describing the physical 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:

$$\boldsymbol{d}(n_t, n_{t-1}) \equiv \begin{bmatrix} Dist(n_t, n_{t-1}) \\ \vdots \\ Dist(n_t, n_{t-1})^3 \end{bmatrix} \quad \boldsymbol{s}_t(n_t, n_{t-1}) \equiv \begin{bmatrix} \ln\left(\frac{S(n_t)}{S(n_{t-1})}\right) \mathbb{1}\left\{\ln\left(\frac{S(n_t)}{S(n_{t-1})}\right) \geq 0\right\} \\ \ln\left(\frac{S(n_t)}{S(n_{t-1})}\right) \mathbb{1}\left\{\ln\left(\frac{S(n_t)}{S(n_{t-1})}\right) < 0\right\} \\ \vdots \\ \ln\left(\frac{S(n_t)}{S(n_{t-1})}\right)^3 \mathbb{1}\left\{\ln\left(\frac{S(n_t)}{S(n_{t-1})}\right) \geq 0\right\} \\ \ln\left(\frac{S(n_t)}{S(n_{t-1})}\right)^3 \mathbb{1}\left\{\ln\left(\frac{S(n_t)}{S(n_{t-1})}\right) < 0\right\} \end{bmatrix}$$

where $Dist(n_t, n_{t-1})$ is the straight-line distance between the centroids of neighborhood n_t and n_{t-1} and $\ln(S(n_t)/S(n_{t-1}))$ is the (natural) log difference in the share of college graduates in neighborhoods n_t and n_{t-1}, with both shares measured at time t. Our measure of social distance captures 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 suggests that 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). We allow this cost to differ depending on "direction" of the move as it may be more costly for low-income renters to move up the distribution of neighborhoods' college shares than down it.

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 1. By Bellman's principle of optimality,[10]

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

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

$$\bar{V}^k\left(x_{it}, \bar{\omega}_t^{kc}\right) \equiv \int V^k\left(s_{it}^c\right) dF^{\varepsilon}(\varepsilon_{it}) \tag{6}$$

[10]The expectation operator $\mathbb{E}_{x'|xn}[\cdot]$ is with respect to the future value of households' observed household-level state variables, x', conditional on households' current state and their neighborhood choice. While the current deterministic setup renders 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.

and define household i's conditional value function as

$$\begin{aligned}(7)\qquad v_n^k\left(x_{it},\bar{\omega}_t^{kc}\right) &\equiv \mathbb{E}_{x'|xn}\left[u_n^k\left(s_{it}^c\right)\right]-\varepsilon_{int}+\delta\mathbb{E}_t\left[\bar{V}^k\left(x_{it+1},\bar{\omega}_{t+1}^{kc}\right)|n,x_{it},\bar{\omega}_t^{kc}\right]\\ &\equiv \bar{u}_n^k\left(x_{it},\bar{\omega}_t^{kc}\right)+\delta\mathbb{E}_t\left[\bar{V}^k\left(x_{it+1},\bar{\omega}_{t+1}^{kc}\right)|n,x_{it},\bar{\omega}_t^{kc}\right]\end{aligned}$$

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

$$(8)\qquad p_n^k\left(x_{it},\bar{\omega}_t^{kc}\right)=\frac{\exp\left(v_n^k\left(x_{it},\bar{\omega}_t^{kc}\right)\right)}{\sum\limits_{n'\in\mathcal{N}^c}\exp\left(v_{n'}^k\left(x_{it},\bar{\omega}_t^{kc}\right)\right)},$$

and the ex ante value function in 6 has the value

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

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):

$$(9)\qquad \bar{V}^k\left(x_{it},\bar{\omega}_t^{kc}\right)=v_n^k\left(x_{it},\bar{\omega}_t^{kc}\right)-\ln\left(p_n^k\left(x_{it},\bar{\omega}_t^{kc}\right)\right)+\gamma,\qquad \forall n\in\mathcal{N}^c.$$

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

$$(10)\qquad e^{\bar{V}}(x',\bar{\omega}_t^{kc},\bar{\omega}_{t+1}^{kc})\equiv\underbrace{\bar{V}(x',\bar{\omega}_{t+1}^{kc})}_{\text{realized}}-\mathbb{E}_{\bar{\omega}'|\bar{\omega}_t^{kc}}\left[\bar{V}(x',\bar{\omega}')|\bar{\omega}_t^{kc}\right]$$

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. Solving 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 we require them 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.[11] Similarly to other

[11] ECCP estimators are so named given their likeness to Euler equations in models with continuous choice variables (Aguirregabiria and Magesan 2013; Kalouptsidi, Scott, and Souza-Rodrigues 2021).

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, and 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. We estimate the second step using 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 2SLS framework.[12]

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 like Rust (1987) that explicitly solve for households' value functions are infeasible. Second, our focus on gentrification implies an inherently nonstationary environment, making modeling the evolution of neighborhood change conceptually and computationally challenging. As we demonstrate when deriving our moment restrictions, ECCP estimation requires neither the complete specification of households' information sets nor a description of how the city-specific state variables evolve. Third, by yielding moment conditions we can evaluate in a 2SLS framework, we can relate our instrumental variables (detailed in Section 6) to the recent literature on quasi-experimental 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_{it}^c evolve according to a first-order Markov process with a transition distribution that factors as

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

(b) *Utility Normalization*: We normalize the utility offered by the outside option for residents with $\tau = 1$, separately for each city:

$$\tilde{\alpha}_{OO^c}^k + \tilde{\beta}_I^k \ln\left(\bar{I}_{OO^ct}^k\right) + \tilde{\beta}_r^k \log\left(r_{OO^ct}\right) + \tilde{\beta}_A^k \ln\left(\frac{\text{Coll}_{OO^ct}}{\text{Pop}_{OO^ct}}\right) + \tilde{\xi}_{OO^ct}^k = \alpha^{kc} \quad \forall\, t$$

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

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

[12]ECCP estimation has been used in a variety of applied settings, from choices over agricultural land use (Scott 2013) and new technology adoption (Groote and Verboven 2019) to 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-Iino 2022). See Kalouptsidi, Scott, and Souza-Rodrigues (2021) for a comprehensive econometric treatment of linear regression techniques with ECCP estimators.

where $e^{\bar{V}}(x', \bar{\omega}_t^{kc}, \bar{\omega}_{t+1}^{kc})$ are the expectational errors defined in equation 10.

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 its own neighborhood choice (cf. Assumption (1) in Kalouptsidi, Scott, and Souza-Rodrigues (2021)). Given that the typical 2010-delineated US census tract contains around 4,000 residents, we believe that this assumption is plausible.[13] Note that Assumption (a) does not require the observed and unobserved city-specific state variables to evolve independently. We highlight this to foreshadow the econometric challenge we face when attempting to identify preferences over functions of ω_t^{kc}.

Assumption (b) says that residents who choose to reside outside of their respective city's urban core obtain a time-invariant and city-specific mean utility.[14] This assumption normalizes each city's neighborhood mean utilities to a constant and time-invariant level, which is necessary to compare welfare across households within CBSAs given that logit models identify only differences in mean utilities. The assumption facilitates exposition and, because each α^{kc} is unobserved, highlights the incommensurability of expected welfare across different k-types and cities.

Last, 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 at time t as well (cf. Lemma 1 in Kalouptsidi, Scott, and Souza-Rodrigues (2021)). The importance of this corollary will become clear in Section 5.3 when we discuss our choice of IVs 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 that are linear in households' demand parameters. To derive these equations, however, we must first introduce the concept of renewal actions.

Renewal Actions. To derive our estimating equations, we use what are termed renewal actions in the dynamic discrete choice literature (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 its current residential tenure. Moreover, because the city-specific state variables, $\bar{\omega}_{t+1}^{kc}$, and unobserved idiosyncratic preference shocks, ε_{it+1}, are independent of the household's state in period t, *all* the remaining state variables reset to a common value upon moving

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

[14] Recall that *all* residents of a given city (i.e., including those outside the urban core) additionally receive a time-varying but neighborhood-invariant utility shock, $\tilde{\alpha}_t^k$. The value of the outside option can, therefore, shift over time, albeit always in proportion to the mean utilities in the urban cores.

to a new neighborhood.[15]

Following the ECCP literature, we exploit such renewal actions when deriving our estimating equation (Scott 2013; Diamond, McQuade, and Qian 2018; Davis et al. 2021; Almagro and Domínguez-Iino 2022). To see how, consider Figure 6. The figure shows the residential choices of two hypothetical type-k households (A and B) with initial tenure length τ between periods $t-1$ and $t+1$. In period $t-1$, these households reside in the same neighborhood. In period t, however, household A moves to neighborhood n while household B leaves the urban core, exercising their outside option OO^c. In period $t+1$, both households move to the same neighborhood, $\bar{n}$. Our estimation procedure involves relating the difference in the expected discounted utilities of the two households' hypothetical neighborhood choices to the difference in the probability that the two households actually make these neighborhood choices. Critically, because moving to neighborhood $\bar{n}$ in period $t+1$ constitutes the same renewal action for both households, their state variables are reset to a common value, equalizing their continuation values. Differences in the expected discounted utilities associated with the two sets of neighborhood choices are thus a function only of households' flow utilities. Constructing estimating equations based on these comparisons therefore helps isolate observed variation in current neighborhood characteristics from variation in households' unobserved continuation values. Throughout the paper we term these consecutive residential location choices *residential paths*.

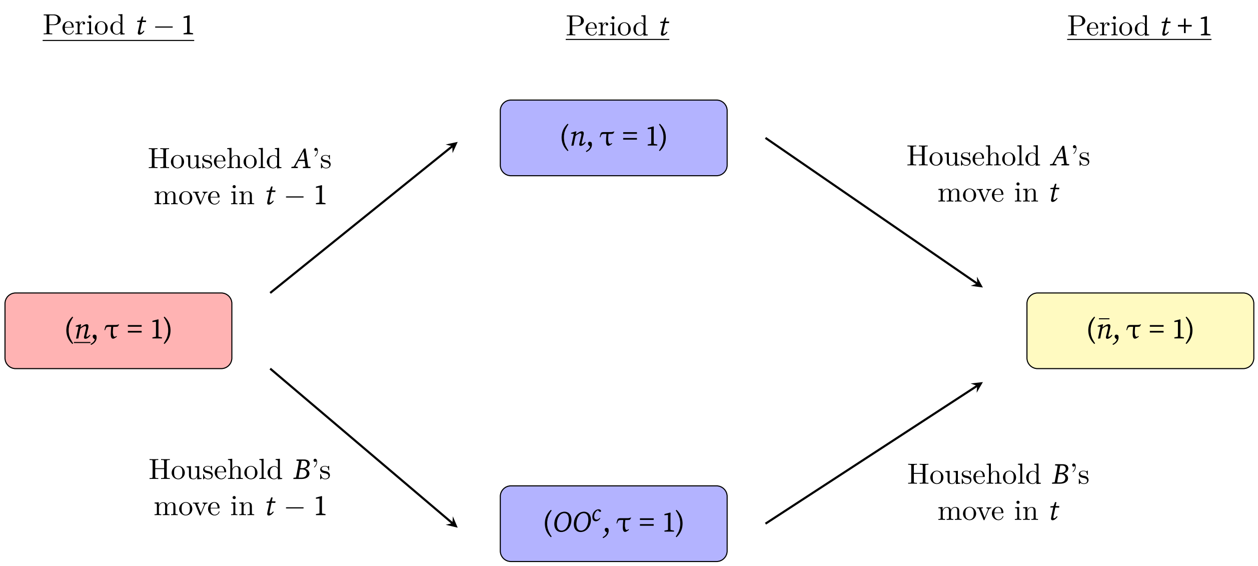


Figure 6. Residential Paths for Two Hypothetical Households

Note: Figure 6 shows hypothetical residential paths for two households, A and B. These households both start in neighborhood $\underline{n}$ with tenure $\tau = 1$, move to different neighborhoods for period t, and convene in a common neighborhood in period $t+1$.

[15]Note that renewal actions depend on our construction of households' neighborhood tenure in Section 4.3. This construction assumes that the lengths of households' prior residential tenures do not impact the value of future residential tenure. While this is a strong assumption, it is necessary to keep the dimension of the household-level state space manageable. We outline the steps we take to mitigate the restrictiveness of this assumption in Appendix B.5.

Our Estimating Equation. Given the residential paths outlined in Figure 6, we can derive the following estimating equation by relating differences in expected discounted utilities associated with the residential paths to the probabilities households take these paths:

$$(11) \qquad Y^k_{x\underline{n}n\bar{n}t} = \bar{\tilde{\alpha}}^k_n + \tilde{\alpha}^k_t + \tilde{\beta}^k_w \ln(\bar{I}^k_{nt}) + \tilde{\beta}^k_r \ln(r_{\tau nt}) + \tilde{\beta}^k_A \ln\left(\frac{\text{Coll}_{nt}}{\text{Pop}_{nt}}\right) + \tilde{\beta}^k_\tau \tilde{\tau}_{xt} + \widetilde{MC}^{kc}_t + \nu^k_{x\underline{n}n\bar{n}t}$$

where

$$\begin{aligned}
Y^k_{x\underline{n}n\bar{n}t} &\equiv \ln\left(\frac{p^k_{xnt}}{p^k_{xOO^ct}}\right) + \delta\left(\sum_{x'} \ln\left(p^k_{x'\bar{n}t}\right) f^x\left(x'|n, x_t\right) - \sum_{x'} \ln\left(p^k_{x'\bar{n}t}\right) f^x\left(x'|OO^c, x_t\right)\right) \\
\tilde{\tau}_{xt} &\equiv \sum_{x'} \ln(\tau_{xt}(x')) f^x\left(x'|n, x_{xt}\right) - \sum_{x'} \ln(\tau_{xt}(x')) f^x\left(x'|OO^c, x_t\right) \\
\widetilde{MC}^{kc}_t &\equiv MC^k_t(n, \underline{n}) - MC^k_t(OO^c, \underline{n}) + \delta\left(MC^k_t(\bar{n}, n) - MC^k_t(\bar{n}, OO^c)\right) \\
\nu^k_{x\underline{n}n\bar{n}t} &\equiv \delta \cdot \tilde{e}(x_t, \bar{\omega}^{kc}_t, \bar{\omega}^{kc}_{t+1}) + \tilde{\xi}^k_{nt}.
\end{aligned}$$

Since the full derivation of this estimating equation is becoming well known, we relegate it to Appendix B.4. 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 11 using linear regression techniques. The following subsection details our two-step estimation procedure.[16]

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.

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.[17] Specifically, we aggregate tenure into two buckets:

$$\bar{\tau} = \begin{cases} 1 & \text{if } \tau \leq 5 \\ 2 & \text{otherwise} \end{cases}$$

[16] Figure 6 illustrates a representative comparison of two residential paths. In practice, to help identify the components of households' fixed moving costs and their preferences for residential tenure, we further consider alternative comparisons of residential paths between neighborhoods and their city's outside option. In Appendix B.5, we present the full set of residential paths used in estimation, discuss how each pair of residential paths contributes to identification, outline all restrictions imposed on the choices of $\underline{n}$, n, and $\bar{n}$, and detail how we compute migration flows between neighborhoods and their outside options.

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

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

$$f^x(x(n_t, \bar{\tau}_t = 2)|n_t, x(n_{t-1}, \bar{\tau}_{t-1})) = \begin{cases} 1, & \text{if } n_t = n_{t-1} \text{ and } \bar{\tau}_{t-1} = 2 \\ g^k_{n_t}, & \text{if } n_t = n_{t-1},\ \bar{\tau}_{t-1} = 1, \text{ and } i \in k \\ 0, & \text{otherwise,} \end{cases}$$

where we estimate $g^k_{n_t}$ directly from the data:

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

$\hat{g}^k_{n_t}$ captures the probability that a household with tenure $\bar{\tau} = 1$ staying in the same neighborhood transitions to $\bar{\tau} = 2$ in the following period. Given that our analysis is at the census tract level, $\hat{g}^k_{n_t}$ is a sparse approximation of $g^k_{n_t}$. This is not problematic for us, however, as we restrict attention to paths originating in tracts with many low-income renters. Appendix B.5 outlines the restrictions we impose on our residential path choices.

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 that a type-k household with state x_{it} moves to each neighborhood $n_t \in \mathcal{N}^c$.[18] 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 that our setting yields particularly sparse empirical choice probabilities, we choose this latter option.

We model the count of k-type households in neighborhood $\underline{n} \in \mathcal{N}^c$ choosing neighborhood $n \in \mathcal{N}^c$ between periods t and $t+1$ as being derived from a Poisson distribution.[19] We parameterize the mean of the Poisson distribution in a way that imposes no additional restrictions on the data-generating process implied by our dynamic model. Specifically, we estimate the following flexible Poisson regression separately for each k-type household:[20]

$$\log\left(\mathbb{E}\left[\left|\underline{n}^k_{\bar{\tau}} \rightarrow n^k_{\bar{\tau}}\right|\right]\right) = \underbrace{\gamma^k_{nt\bar{\tau}=1} \cdot \mathbb{1}\{\bar{\tau} = 1, n = \underline{n}\} + \gamma^k_{nt\bar{\tau}=2} \cdot \mathbb{1}\{\bar{\tau} = 2, n = \underline{n}\}}_{\text{Mean neighborhood utilities and CVs for stayers}} \tag{12}$$

[18] We could similarly employ any nonparametric method to compute $\hat{p}^k_{xnt}$ directly from the data.

[19] We choose to model the data as a Poisson distribution because of its ability to account for sparse data, its robustness to the inclusion of many fixed effects ?, and its computational efficiency (Correia, Guimarães, and Zylkin 2020).

[20] 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.3 provides details our Poisson regressions and shows how our regression specification does not impose additional restrictions on the neighborhood choice problem of households in our dynamic model.

$$+ \underbrace{\mu_{nt}^{k} \cdot \mathbb{1}\{n \neq \underline{n}\} + \lambda^{kc} \cdot \mathbb{1}\{\underline{n} = OO^c, n \neq \underline{n}\}}_{\substack{\text{Mean neighborhood utilities, CVs,} \\ \text{and fixed moving costs for movers}}}$$

$$+ \underbrace{\left(\beta_d^{k\prime}\, \boldsymbol{d}(n, \underline{n}) + \beta_s^{k\prime}\, \boldsymbol{s}(n, \underline{n})\right) \cdot \mathbb{1}\{n \neq OO^c, \underline{n} \neq OO^c\}}_{\text{Variable moving costs}}$$

where $\left|\underline{n}_{\bar{\tau}}^{k} \rightarrow n_{\bar{\tau}}^{k}\right|$ is the count of k-type households with aggregate tenure status $\bar{\tau}$ in neighborhood $\underline{n}$ choosing neighborhood n in period t. $\gamma_{nt\bar{\tau}}^{k}$ are tenure- and period-specific fixed effects capturing the value to households of staying in their current neighborhood. μ_{nt}^{k} is a fixed effect capturing the mean value associated with moving to neighborhood n, and λ^{kc} captures the fixed cost from moving into a city's urban core from its outside option. Finally, $\beta_d^{k\prime}\, \boldsymbol{d}(n, \underline{n})$ and $\beta_s^{k\prime}\, \boldsymbol{s}(n, \underline{n})$ are the variable moving costs specified in Section 4.3.

We estimate this Poisson model separately for each k-type, but pool observations across periods, aggregate tenures, and cities. We use the estimates from the model to predict the probability a k-type household with aggregate tenure status $\bar{\tau}$ living in neighborhood $\underline{n}$ chooses neighborhood n in each year: $\hat{p}_{xnt}^{k}$.[21] Equation 12 additionally identifies our variable moving cost parameters, $\beta_d^{k\prime}$ and $\beta_s^{k\prime}$. Since the cost of moving to neighborhood n differs for each k-type household depending on its origin neighborhood, we can separately identify the parameters governing variable moving costs from $\gamma_{nt\bar{\tau}}^{k}$, μ_{nt}^{k}, and λ^{kc}. Variation in the distance households move within their urban core, conditional on moving, identifies $\beta_d^{k\prime}$. Similar variation for social distance identifies the cost of moving to neighborhoods socially different to one's origin neighborhood, $\beta_s^{k\prime}$.[22] We report the full set of variable moving cost estimates in Appendix Table **??**.

Step Two. Given our estimated variable moving cost parameters $\left(\hat{\beta}_s^{k\prime} \text{ and } \hat{\beta}_d^{k\prime}\right)$, our estimated conditional choice probabilities $\left(\hat{p}_{xnt}^{k}\right)$, and our estimated transition probabilities $\left(\hat{f}^{x}\right)$ from step 1, we may now construct the empirical analogue of equation 11:

$$\hat{Y}_{x\underline{n}n\bar{n}t}^{k} = \tilde{\bar{\alpha}}_n^{k} + \tilde{\alpha}_t^{k} + \tilde{\beta}_w^{k} \ln(\bar{I}_{nt}^{k}) + \tilde{\beta}_r^{k} \ln(\hat{r}_{\tau nt}) + \tilde{\beta}_A^{k} \ln\left(\frac{\text{Coll}_{nt}}{\text{Pop}_{nt}}\right) + \tilde{\beta}_\tau^{k} \hat{\tilde{\tau}}_{xt} + \widetilde{\widetilde{MC^F}}_t^{kc} + \nu_{x\underline{n}n\bar{n}t}^{k}, \tag{13}$$

where $\widetilde{\widetilde{MC^F}}_t^{kc}$ is the difference in the fixed portion of moving costs as the variable portion is estimated in the first-step and embedded in $\hat{Y}_{x\underline{n}n\bar{n}t}^{k}$. We define Equation 13 fully in Appendix B.4.

To be precise about identification, it is worth unpacking the error term, $\nu_{x\underline{n}n\bar{n}t}^{k}$, in equation 13. $\nu_{x\underline{n}n\bar{n}t}^{k}$ is comprised of both unobserved neighborhood-specific amenities, $\tilde{\xi}_{nt}^{kc}$, and expectational errors,

[21] Many census tracts lack k-type current residents 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.

[22] Note that our Poisson distribution assumption in equation 12 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 (represented in equation 8) for these continuous variables, yielding identical maximum likelihood estimation (MLE) estimates (Guimarães, Figueirdo, and Woodward 2003).

$\tilde{e}(x_t, \bar{\omega}_t^{kc}, \bar{\omega}_{t+1}^{kc})$.[23] We consider each of these in turn starting with the unobserved neighborhood-specific amenities, $\tilde{\xi}_{nt}^{kc}$. Since we do not restrict the relationship between $\tilde{\xi}_{nt}^{kc}$ and the remaining time-varying observable neighborhood characteristics, ordinary least squares (OLS) estimates of 13 would be biased. Expected income, neighborhood-level housing costs, and the share of college graduates are invariably correlated with many unobserved neighborhood-level factors, such as proximity to natural amenities that we do not observe as econometricians (Lee and Lin 2018).

To distinguish between preferences for observed versus unobserved neighborhood amenities, we start by differencing equation 13 using residential paths starting in 2017 (i.e., $t-1=2017$) and residential paths starting in 2010 (i.e., $t-1=2010$), obtaining,

$$\Delta \hat{Y}_{x\underline{n}n\bar{n}}^{k} = \Delta\tilde{\alpha}_t^k + \tilde{\beta}_w^k \Delta \ln(\bar{I}_{nt}^k) + \tilde{\beta}_r^k \Delta \ln(\hat{r}_{\tau n}) + \tilde{\beta}_A^k \Delta \ln\left(\frac{\text{Coll}_n}{\text{Pop}_n}\right) + \tilde{\beta}_\tau^k \Delta\hat{\tilde{\tau}}_x + \Delta\nu_{x\underline{n}n\bar{n}}^k \tag{14}$$

where Δ corresponds to the difference in the associated variable between $t=2011$ and $t=2018$. Differencing equation 13 removes the time-invariant component of exogenous neighborhood amenities, $\tilde{\alpha}_n^k$, and the time-invariant components of the moving costs variables, MC^{kc}, MC_{OO}^{kc}, and $\hat{\boldsymbol{\beta}}_d^{k\prime}\boldsymbol{d}(n,\underline{n})$.[24] 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:

$$\begin{aligned} 0 &= \mathbb{E}\left[z_n\, \Delta\nu_{x\underline{n}n\bar{n}}^k\right] \\ &= \mathbb{E}\left[z_n\, \left(\Delta\xi_n^k + \Delta\, \left(\delta\cdot\tilde{e}(x, \bar{\omega}_t^{kc}, \bar{\omega}_{t'}^{kc})\right)\right)\right] \\ &= \mathbb{E}\left[z_n\, \left(\Delta\xi_n^k + \Delta\left(\delta\cdot e^{\bar{V}}(n, x, \bar{\omega}_t^{kc}, \bar{\omega}_{t'}^{kc}) - \delta\cdot e^{\bar{V}}(OO^c, x, \bar{\omega}_t^{kc}, \bar{\omega}_{t'}^{kc})\right)\right)\right] \end{aligned} \tag{15}$$

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

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

In addition to being orthogonal to changes in unobserved neighborhood amenities, equation 15 shows that our instruments must also be mean independent of changes in households' expectational errors—the second component of $\nu_{xnn'\bar{n}t}^k$. Recall Assumption (c) 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

[23] Specifically, $\nu_{x\underline{n}n\bar{n}t}^k \equiv \delta\cdot\tilde{e}(x_t, \bar{\omega}_t^{kc}, \bar{\omega}_{t+1}^{kc}) + \tilde{\xi}_{nt}^k$.

[24] We estimate the remaining time-invariant parameters in a final stage. Specifically, we estimate equation 13 conditional on the estimates from our differenced regressions and our first-step Poisson regression. We assume that the effect of residential tenure and moving costs on the likelihood of different residential paths is uncorrelated with unobserved neighborhood amenities.

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 from the information in households' period-t information sets.

To see why our instruments must not be constructed using elements outside households' contemporaneous information sets, consider an instrument that shocks the neighborhood-n elements of $\omega_{t'}^{kc}$ for any $t \leq 2017$.[25] Assume that this shock cannot be predicted with the information in households' contemporaneous information sets, $\mathcal{I}_{it}$. If it is relevant, the instrument will be mechanically 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, $\mathbb{E}_{\bar{\omega}'|\bar{\omega}_t^{kc}}\left[\bar{V}(x', \bar{\omega}')|\bar{\omega}_t^{kc}\right]$, violating the exclusion restriction embodied in 15. For this reason, the instruments we detail immediately below are designed to predict changes in our endogenous variables through 2011–2018 using variation that can be predicted solely from households' 2010 information sets.

6. Identification

We estimate Equation 14 via two-stage least squares (2SLS). We construct three instruments to exactly identify the endogenous variables: $\tilde{\beta}_w^k$, $\tilde{\beta}_r^k$, and $\tilde{\beta}_A^k$. Our first instrument interacts neighborhood-level measures of proximity to relevant employment opportunities with nationwide trends in tradable-industry employment. This is our "job market access" instrument. It helps identify $\tilde{\beta}_w^k$ by shifting neighborhoods' expected commute-time-discounted income, $\bar{I}_{nt}^k$. Our second instrument predicts changes in a neighborhood's share of college graduates by interacting measures of proximity to other neighborhoods' shares of college graduates with growing CBSA-wide demand among this population. This is our "proximity" instrument and it is particularly helpful in identifying households' preferences for endogenous amenities, $\tilde{\beta}_A^k$. By simultaneously shifting access to non-tradable employment opportunities and therefore expected income, $\bar{I}_{nt}^k$, it also helps identify $\tilde{\beta}_w^k$. Our third instrument interacts the proximity instrument with tracts' baseline urban development levels, helping identify households' disutility from paying higher rents, $\tilde{\beta}_r^k$. Neighborhoods' baseline urban development levels proxy for their housing supply elasticities, inducing variation in rent conditional on changes in neighborhood demand.

When estimating Equation 14, we include a similar set of baseline destination-tract controls, notated as $X_{nt=2010}$, as we did in Section XX.[26]. These controls ensure identification comes from how households' migration choices change when endogenous neighborhood characteristics shift among tracts that are observably similar in 2010—our baseline year. The evolution of exogenous neighborhood amenities may therefore vary systematically with our instruments across tracts with different baseline characteristics

[25] 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 shows that we must construct z_n using variation that can be predicted from households' 2010 information sets.

[26] The tract-level controls include quadratic polynomials for each continuous variable: population-weighted distance to the Metropolitan Division's CBD, log college-graduate share, log hedonic rent, white and Black population shares, and urban land cover. We also include ten indicators for equally sized population-weighted distance rings centered at each Metropolitan Division's CBD. We take logs of college shares and hedonic rents to match the scale of the baseline endogenous variables. Relative to our descriptive analyses, we replace the residential market access controls with a suite of job market access measures that support identification through shocks to national industry growth (Borusyak, Hull, and Jaravel 2022). We discuss these controls below and in Appendix C.

while still satisfying the augmented moment restriction:

$$0 = \mathbb{E}\left[z_n\, \Delta\nu^k_{x\underline{n}n\bar{n}} \middle| X_{nt=2010}\right].$$

We further include origin-tract fixed effects, notated as $\delta_{\underline{n}}$, when estimating Equation 14. These fixed effects ensure that identifying variation in households' migration choices comes from residential paths originating from the same neighborhood $\underline{n}$. This is important given the potential for non-random changes in origin tracts' exogenous amenities $\left(\Delta\xi^k_{\underline{n}}\right)$ to effect neighborhood exit rates and thus overall move probabilities to nearby tracts, whose characteristics may be correlated with $\Delta\xi^k_{\underline{n}}$. For example, if origin tracts with declining exogenous amenities are located near neighborhoods with low measures of proximity to other neighborhoods' shares of college graduates—which households in neighborhood $\underline{n}$ had a high ex-ante probability of moving to—our proximity instrument would not be exogenous. The final augmented moment restriction becomes,

$$0 = \mathbb{E}\left[z_n\, \Delta\nu^k_{x\underline{n}n\bar{n}} \middle| X_{nt=2010}, \delta_{\underline{n}}\right],$$

and it is this equation we assess the validity of our instruments against.

Job Market Access IV. We predict changes in neighborhoods' expected commute-time-discounted income $\left(\bar{I}^k_{nt}\right)$ by interacting baseline measures of proximity to relevant employment opportunities with nationwide trends in tradable industry employment. We term the measures of proximity to relevant employment opportunities as "job market access" measures. We build our instrumental variable from neighborhood- and industry-specific measures of job market access for both k-type households:

$$\text{JMA}^k_{n\iota t} = \sum_{m \in \mathcal{N}^c \setminus n} e^{-\eta^{ck}\tau^k_{n,m}} l^k_{m\iota t}, \tag{16}$$

where $l^k_{m\iota t}$ is the number of jobs held by low-income workers of race k in workplace tract m, 6-digit NAICS industry ι, and time t; where $\tau^k_{n,m}$ is the type-specific travel time between tracts n and m; and where η^{ck} is a CBSA- and type-specific spatial decay parameter governing the importance of faraway jobs relative to closer ones for quantifying job market access.[27] Like $\bar{I}^k_{nt}$, we derive Equation 16 from the household workplace choice problem and describe its empirical construction in Appendix B.1.

We interact our industry-specific measures of job market access with changes in national industry employment:

$$\Delta\widetilde{\text{JMA}}^k_{nt_0t} = \sum_{\iota \in \mathcal{T}} \frac{\text{JMA}^k_{n\iota t_0}}{\sum_{\iota'} \text{JMA}^k_{n\iota' t_0}} \cdot \ln\left(\frac{L^c_{\iota t}}{L^c_{\iota t_0}}\right), \tag{17}$$

where $L^c_{\iota t}$ is national industry employment in period t and industry ι excluding that employment in CBSA c which we obtain from the Longitudinal Business Database (LBD). $\mathcal{T}$ denotes the set of tradable industries we use to predict changes in expected commute-time-discounted income. We obtain measures of industry trade costs from Gervais and Jensen (2019) and use them to exclude non-tradable industries

[27]We use all low-income workers of race k—not just renters—when constructing these measures.

from $\Delta\widetilde{\text{JMA}}^{k}_{nt_0t}$.[28] Recall that, to satisfy the exclusion restriction embodied in 15, households must predict national employment growth using only the information available in period t_0. Choosing t and t_0 in 17 therefore involves a distinct trade-off: moving the employment growth horizon closer to our analysis period strengthens the first-stage but simultaneously assumes that households can accurately predict employment growth further into the future. We set $t_0 = 2010$ and $t = 2013$ because this is the shortest horizon that delivers sufficient first-stage power. Our instrument construction therefore implicitly exploits serial correlation in industry employment growth between 2010-2013 and 2013-2019.

Our job market access instrument builds on Severen (2023), Baum-Snow, Hartley, and Lee (2019), and Baum-Snow and Han (2023). In contrast to these papers, we disaggregate job market access at the 6-digit NAICS level rather than at 2-digit sectors reported in the public LEHD Origin-Destination Employment Statistics. This finer disaggregation serves two purposes. First, by focusing on tradable employment growth within each sector, we preserve the Bartik shift-share logic and avoid evolving consumer preferences jointly influencing households' location choices and industry employment (Bartik 1991). For example, evolving preferences for different types of nontradable services may simultaneously influence employment in those industries and households' within-CBSA residential location choices (Couture and Handbury 2020). Second, substantial variation in 6-digit NAICS industry employment growth during 2010-2013 lets us treat national industry growth—weighted by each tract's 2010 employment mix—as conditionally exogenous while allowing initial establishment locations to correlate with changing exogenous amenities (Borusyak, Hull, and Jaravel 2022).

Intuitively, identification based on our job market access instrument (17) requires that (i) households can predict national industry employment growth through 2010-2013 using only information available to them in 2010 and (ii) establishments in industries with increasing or decreasing nationwide employment are not concentrated near tracts with systematically changing exogenous amenities, conditional on our rich set of tract-level controls and origin-tract fixed effects. Following Borusyak, Hull, and Jaravel (2022), we further relax (ii) by conditioning on baseline exposure to 2-digit sector employment, so only tract-weighted within-sector residual employment growth must be orthogonal to changing exogenous neighborhood amenities. We discuss these and other identification considerations thoroughly and more formally in Appendix C.

Proximity IV. We predict changes in a neighborhood's college-graduate share by interacting its distance-weighted exposure to other neighborhoods' college-graduate shares with changes in CBSA-wide demand among this population. The instrument is motivated by research showing that, among low-income neighborhoods, those closest to high-income neighborhoods experience the greatest home price appreciation during positive labor demand shocks (Guerrieri, Hartley, and Hurst 2013)—a proxy for gentrification. We construct the proximity instrument using the geographic-diffusion model in Saiz and Wachter (2011), closely resembling the gentrification instruments of Glaeser, Luca, and Moszkowski (2023) and Brummet

[28] We find that excluding industries among the highest 25% of trade costs in Gervais and Jensen (2019) maximizes first-stage power. We crosswalk Gervais and Jensen (2019)'s 2007 NAICS codes to 2012 NAICS codes using industry employment weights obtained from the 2010 LBD cross section which records both 2007 and 2012 NAICS codes for each establishment. For 2007-defined industries that are assigned to a single 2012-defined NAICS code or for industries that amalgamate, we construct an employment-weighted average of the 2007-defined trade costs.

and Reed (2021). Specifically, we construct,

$$\widetilde{\Delta \mathrm{Prox}}_{nt_0t} = \sum_{m \in \mathcal{N}^c \setminus n} e^{-\rho \tau_{n,m}} \frac{\mathrm{Coll}_{m,t_0}}{\mathrm{Pop}_{m,t_0}} \cdot \ln \left(\frac{\mathrm{Coll}^c_t}{\mathrm{Coll}^c_{t_0}} \right), \tag{18}$$

where $\tau_{n,m}$ is the straight-line distance between census tracts n and m, and ρ is a spatial decay parameter governing the importance of faraway neighborhoods' college graduate shares relative to closer neighborhoods' shares. As $\rho \to \infty$, only the neighborhoods closest to n matter for determining $\widetilde{\Delta \mathrm{Prox}}_{nt_0t}$. Conversely, as $\rho \to 0$, every neighborhood receives the same weight in $\widetilde{\Delta \mathrm{Prox}}_{nt_0t}$, eliminating any within-CBSA variation and rendering the instrument irrelevant. Since we do not have a good prior for ρ, we calibrate its value via k-fold cross-validation, choosing ρ that minimizes the out-of-sample sum of squared errors in a set of unweighted tract-level first-stage regressions. Multiplying our proximity measures by log-changes in CBSAs' college graduate population, $\ln \left(\frac{\mathrm{Coll}^c_t}{\mathrm{Coll}^c_{t_0}} \right)$, helps ensure instrument monotonicity. We again select $t_0 = 2010$ and $t = 2013$ for our instrument horizon.

Identification based on our proximity instrument (18) requires that (i) households can predict[29] their CBSA's growth in college graduates over 2010-2013 using only information available in 2010 and (ii) a neighborhood's proximity-weighted exposure to other neighborhoods' college-graduate shares is uncorrelated with changes in exogenous amenities once we condition on our rich set of tract-level controls. This latter condition permits changes in neighborhood college shares to alter the local mix of private consumption amenities (Almagro and Domínguez-Iino 2022) or the supply of public amenities like crime (Porreca 2023) or school quality (Pearman 2018). Indeed, β^k_A captures both households' direct preference for a neighborhood's college-graduate share and their preference for the way college graduates reshape local private and public amenities, following Diamond (2016) and Su (2022). We assume, however, that conditional on observable neighborhood characteristics, exogenous amenity shocks that alter college-graduate shares are uncorrelated with $\widetilde{\Delta \mathrm{Prox}}_{nt_0t}$. For instance, although we recompute market-access measures each year, if public infrastructure investments both correlate with $\widetilde{\Delta \mathrm{Prox}}_{nt_0t}$ and raise neighborhood college-graduate shares, we cannot disentangle low-income renters' preference for improved transit access from preferences for endogenous amenities $\left(\tilde{\beta}^k_A\right)$.

We predict rent changes by interacting the proximity instrument with baseline tract-level urban development measures from Baum-Snow and Han (2023).[30] Because these development measures also enter Equation 14 directly as controls ($X_{nt=2010}$), identification proceeds analogously to a difference-in-difference estimator. We effectively compare differences in rent growth between heavily and lightly developed tracts when $\widetilde{\Delta \mathrm{Prox}}_{nt_0t}$ is low to corresponding differences when $\widetilde{\Delta \mathrm{Prox}}_{nt_0t}$ is high. This setup mirrors the price instruments in Diamond (2016) and permits changes in exogenous amenities to covary with the development measures (Davidoff 2016).

[29] Predict is doing a lot of work; they must have rational expectations.

[30] We use the share of urban land development in the census tract in 2011. Baum-Snow and Han (2023) suggest researchers use housing supply elasticities generated from their quadratic finite mixture models which are functions of, among other tract characteristics, urban land development. As such, the two measures are highly correlated within CBSAs. In our data, and after conditioning on a rich set of tract-level controls, the raw urban development shares have a stronger first-stage relationship with subsequent rent growth than the estimated elasticities do, a pattern consistent with attenuation from estimation uncertainty in the latter.

7. Parameter Estimates

Discussion coming soon.

TABLE 2. Parameter Estimates

	(1)		(2)		(3)	
	Black	Non-Black	Black	Non-Black	Black	Non-Black
Rent $\left(\tilde{\beta}_r^k\right)$	-4.105	-3.476	-2.074	-2.405	-2.633	-3.133
	(0.327)	(0.152)	(0.355)	(0.213)	(0.383)	(0.280)
College share $\left(\tilde{\beta}_A^k\right)$	2.206	2.206	—	—	0.485	1.293
	(0.102)	(0.127)			(0.253)	(0.390)
	[$317]	[$415]			[$107]	[$267]
Expected income $\left(\tilde{\beta}_w^k\right)$	—	—	5.197	3.672	4.226	1.681
			(0.343)	(0.314)	(0.684)	(0.684)
			[$1,627]	[$1,042]	[$997]	[$349]
Origin-tract fixed effects	✓	✓	✓	✓	✓	✓
Baseline destination-tract controls	✓	✓	✓	✓	✓	✓
BHJ 2022 exposure controls	✓	✓	✓	✓	✓	✓
SW F-stat (Rent)	102.9	544.8	134.7	324.3	93.69	42.66
SW F-stat (College share)	238.9	466	—	—	78.31	53.67
SW F-stat (Expected income)	—	—	156	148.2	83.26	45.57
Kleibergen-Paap F-stat	67.84	320.9	90.69	78.96	22.86	9.29
Residential paths (000s)	58,320	71,610	58,320	71,610	58,320	71,610

Notes: Table 2 reports 2SLS estimates of the endogenous variables in Equation 14. Column (1) omits neighborhood expected income; Column (2) omits neighborhood college-graduate shares. We describe the residential paths used in estimation and their weights in Appendix B.5. Standard errors, in parentheses, are clustered by origin tract. Values in square brackets translate the coefficients into households' willingness to pay for a 10 percent increase in a tract's expected income or college-graduate share, measured in annual rents. We report the full coefficients for baseline destination-tract controls in Table A9 and describe our preferred market-access exposure controls following Borusyak, Hull, and Jaravel (2022) in the main text and Appendix C. "SW F-stat" denotes the Sanderson–Windmeijer conditional F-statistic.

Robustness. Coming soon.

8. Welfare Analysis

Discussion and estimates coming soon. See appendix for details.

Table 3. Moving Costs

	Black	Non-Black
Panel A: Fixed moving costs and neighborhood tenure		
High neighborhood tenure $\left(\tilde{\beta}^k_\tau\right)$	0.351 (0.073) [$753]	0.255 (0.0329) [$520]
Fixed tract moving cost $\left(MC^{kc}\right)$	-3.319 [$15,240]	-3.221 [$11,942]
Fixed CBSA moving cost $\left(MC^{kc}_{OO}\right)$	-3.368 [$15,639]	-3.348 [$12,711]
Origin-tract fixed effects	✓	✓
Destination-tract fixed effects	✓	✓
Year fixed effects	✓	✓
Residential paths (000s)	423,300	506,400

Panel B: Variable moving costs	Black			Non-Black		
	Linear	Quadratic	Cubic	Linear	Quadratic	Cubic
Physical distance $\left(\beta^k_d\right)$	-.273 (.0000308)	.00589 (1.69e-06)	-.000049 (2.41e-08)	-.319 (0.0000213)	.00741 (1.13e-06)	-.0000611 (1.55e-08)
Upward social distance $\left(\beta^k_{s,\text{Up}}\right)$	.128 (.000518)	-.245 (.000585)	.0289 (.000177)	.0875 (.000358)	-.548 (.000434)	.0951 (.000138)
Downward social distance $\left(\beta^k_{s,\text{Down}}\right)$	-.179 (.000576)	-.218 (.000701)	.029 (.000225)	-.141 (.000373)	-.287 (.000424)	.0479 (.000129)
Pseudo R-squared		.9438			.9555	
PIK-Years (000s)		9,472			25,470	

Notes: Panel A reports the high neighborhood tenure coefficients and pseudo-medians of the city-specific fixed moving costs estimated from Equation 13, conditional on the coefficient estimates reported in column (3) of Table 2. We report additional pseudo-percentiles of the city-specific fixed moving costs in Table **??**. We pool all cross-sections from 2010–2019 and use the same residential paths and weights as when estimating Equation 14. Standard errors, in parentheses, are clustered by origin-tract. Values in square brackets translate the coefficients into households' willingness to pay to avoid leaving their origin tract, and show how this willingness to pay rises for households with high aggregate neighborhood tenure, measured in annual rents. Panel B reports the variable moving cost coefficients estimated from Equation 12. We weight observations for each origin-tract, year, and tenure group by the number of sample households in that cell, while each outside-option cell is assigned the maximum weight of any observation within the same CBSA, year, and tenure group. We smooth raw move counts using a three-year rolling window, as detailed in Appendix B.3. We report the total number of PIK-years included in the estimation sample. We translate the variable moving cost coefficient estimates to willingness to pay measures in Figure A4.

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 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.

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.

Ahlfeldt, Gabriel M, Stephen J Redding, Daniel M Sturm, and Nikolaus Wolf. 2015. "The Economics of Density: Evidence From the Berlin Wall." *Econometrica* 83 (6): 2127–2189.

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

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

Bartik, Timothy. 1991. *Who Benefits from State and Local Economic Development Policies?.*: W.E. Upjohn Institute.

Baum-Snow, Nathaniel, and Giles Duranton. 2025. "Housing supply and housing affordability." *Handbook of Regional and Urban Economics* 5: 353–461.

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.

Beauregard, Pierre-Loup. 2024. "Gentrification, displacement, and income trajectory of incumbents." *Working Paper.*

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.*

Bond, Brittany, J. David Brown, Adela Luque, and Amy O'Hara. 2014. "The Nature of the Bias When Studying Only Linkable Person Records: Evidence frmo the American Community Survey." *CARRA Working Paper Series* (2014-08).

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

Brummet, Quentin. 2014. "Comparison of Survey, Federal, and Commercial Address Data Quality." *CARRA Working Paper* (2014-06).

Brummet, Quentin, Diane Cronkite, Denise Flanagan-Doyle, and Kevin Rinz. 2016. "Analysis Results: A Comparison of Consumer Expenditure Survey Data with CoreLogic Property Tax and Deeds Data." *CARRA Working Paper.*

Brummet, Quentin, and Davin Reed. 2021. "The Effects of Gentrification on Incumbent Residents." *Working Paper.*

Cattaneo, Matias D, Richard K Crump, Max H Farrell, and Yingjie Feng. 2024. "On Binscatter." *American Economic Review* 114 (5): 1488–1514.

Clark, Sandra, Dorothy Barth, Ariel Binder, Adji Fatou Diagne, Nikolas Pharris-Ciurej, John Voorheis, Evan Brassell, and Christopher Mazur. 2023. "Acreage Administrative Records Data Research for the American Community Survey." *2023 AMERICAN COMMUNITY SURVEY RESEARCH AND EVALUATION REPORT MEMORANDUM SERIES* ACS23-RER-16.

Collinson, Robert, John Eric Humphries, Nicholas Mader, Davin Reed, Daniel Tannenbaum, and Winnie van Dijk. 2024. "Eviction and Poverty in American Cities." *Quarterly Journal of Economics* 139: 57–120.

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.

Davidoff, Thomas. 2016. "Supply Constraints Are Not Valid Instrumental Variables for Home Prices Because They Are Correlated With Many Demand Factors." *Critical Finance Review* 5: 177–206.

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.

DeLuca, Stefanie, Holly Wood, and Peter Rosenblatt. 2019. "Why poor families move (and where they go): Reactive mobility and residential decisions." *City & Community* 18 (2): 556–593.

Desmond, Matthew. 2017. *Evicted: Poverty and profit in the American city.*: Crown.

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, Tim McQuade, and Franklin Qian. 2019. "The effects of rent control expansion on tenants, landlords, and inequality: Evidence from San Francisco." *American Economic Review* 109 (9): 3365–3394.

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, 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.

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 M O'Regan. 2011. "How Low Income Neighborhoods Change: Entry, Exit, and Enhancement." *Regional Science and Urban Economics* 41 (2): 89–97.

Ferreira, Fernando V, Jeanna H Kenney, and Benjamin Smith. 2023. "Household Mobility, Networks, and Gentrification of Minority Neighborhoods in the US."Technical report, National Bureau of Economic Research.

Freeman, Lance. 2005. "Displacement or Succession? Residential Mobility in Gentrifying Nneighborhoods." *Urban Affairs Review* 40 (4): 463–491.

Freeman, Lance. 2006. *There goes the hood: Views of gentrification from the ground up.*: Temple University Press.

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.

Gallin, Joshua, Lara Loewenstein, Hugh Montag, and Randal Verbrugge. 2024. "Sticky Continuing-Tenant Rents." *Working Paper.*

Gans, Herbert J. 1982. *Urban Villagers.*: Simon and Schuster.

Garin, Andrew, Ethan Jenkins, Evan Mast, and Bryan A Stuart. 2025. "Dynamic Individuals, Static Neighborhoods: Migration and Earnings Changes in Poor Neighborhoods." *Working Paper.*

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, Michael Luca, and Erica 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.

Graham, Matthew, Mark Kutzbach, and Danielle Sandler. 2017. “Developing a Residence Candidate File for Use With Employer-Employee Matched Data.” *CES Working Paper Series.*

Graham, Matthew, Erika McEntarfer, Kevin McKinney, Stephen Tibbets, and Lee Tucker. 2022. “LEHD Snapshot Documentation, Release S2021-R2022Q4.” *CES Working Paper Series.*

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.

Gu, Yizhen, Naijia Guo, Jing Wu, and Ben Zou. 2021. “Home Location Choices and the Gender Commute Gap.” *Journal of Human Resources.*

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.

Hanson, Gordon H. 2005. “Market potential, increasing returns and geographic concentration.” *Journal of International Economics* 67 (1): 1–24.

Harris, Chauncy D. 1954. “The Market as a Factor in the Localization of Industry in the United States.” *Annals of the Association of American Geographers* 44 (4): 315–348.

Herzog, Ian. 2021. “National Transportation Networks, Market Access, and Regional Economic Growth.” *Journal of Urban Economics* 122: 103316.

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.

Jargowsky, Paul A. 2009. “Immigrants and Neighbourhoods of Concentrated Poverty: Assimilation or Stagnation?” *Journal of Ethnic and Migration Studies* 35 (7): 1129–1151.

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.

Kan, Kamhon. 2007. “Residential mobility and social capital.” *Journal of Urban Economics* 61 (3): 436–457.

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

Koenen, Martin, and Drew Johnston. 2024. “Social Ties and Residential Choice: Micro Evidence and Equilibrium Implications.” *Working Paper.*

Lee, Sanghoon, and Jeffrey Lin. 2018. “Natural Amenities, Neighbourhood Dynamics, and Persistence in the Spatial Distribution of Income.” *The Review of Economic Studies* 85 (1): 663–694.

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..”

McKinnish, Terra, Randall Walsh, and T Kirk White. 2010. “Who gentrifies low-income neighborhoods?” *Journal of urban economics* 67 (2): 180–193.

Meltzer, Rachel, and Pooya Ghorbani. 2017. “Does Gentrification Increase Employment Opportunities in Low-Income Neighborhoods?” *Regional Science and Urban Economics* 66: 52–73.

Pattillo, Mary. 2010. *Black on the block: The politics of race and class in the city.*: University of Chicago Press.

Pearman, Francis. 2018. “Gentrification and Academic Achievement: A Review of Recent Research.” *Review of Educational Research* 89 (1): 125–165.

Pennington, Kate. 2021. “Does Building New Housing Cause Displacement?: The Supply and Demand Effects of Construction in San Francisco.” *Working Paper.*

Porreca, Zachery. 2023. “Gentrification, gun violence, and drug markets.” *Journal of Economic Behavior & Organization* 207: 235–256.

Raphael, Steven, and Lorien Rice. 2002. “Car ownership, employment, and earnings.” *Journal of Urban Economics* 52: 109–130.

Rust, John. 1987. “Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher.” *Econometrica*

55 (5): 999–1033.

Saiz, Albert, and Susan Wachter. 2011. “Immigration and the Neighborhood.” *American Economic Journal: Economic Policy* 3 (2): 169–88.

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

Severen, Christopher. 2023. “Commuting, Labor, and Housing Market Effects of Mass Transportation: Welfare and Identification.” *The Review of Economics and statistics* 105 (5): 1073–1091.

Steven Ruggles, Matthew Sobek Daniel Backman Grace Cooper Julia A. Rivera Drew Stephanie Richards Renae Rodgers Jonathan Schroeder, Sarah Flood, and Kari C.W. Williams. “IPUMS USA: Version 16.0 [dataset].”

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

Sullivan, John, and Katie Genadek. 2024. “Using the Census Bureau’s Master Address File for Migration Research.” *ADEP Working Paper Series.*

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.

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.*

Warnes, Pablo. 2024. “Transport Infrastructure Improvements and Spatial Sorting: Evidence from Buenos Aires.” *Unpublished manuscript.*

Appendix A. Sample and Variable Construction

A.1. Data Sources

Residential Locations. We record residential location histories using two complementary sources developed by the U.S. Census Bureau: the Master Address File-Auxiliary Reference File (MAF-ARF) and, post 2011, the Longitudinal Employer-Household Dynamics' Individual Characteristics File (LEHD-ICF). Both rely on the Census Bureau's Master Address File (MAF), a continuously updated inventory of all known U.S. housing units that serves as the sample universe for, among other surveys, the Decennial Census and the American Community Survey (ACS). Each address in the MAF is assigned a stable identifier (MAFID) and includes detailed physical and census-geography attributes. Each year, the Census Bureau commits the live MAF into a research extract known as the Master Address File Extract (MAF-X), which contains a preferred address, the address' MAFID, and many of the geographic identifiers and property characteristics for every housing unit (Sullivan and Genadek 2024; Graham, Kutzbach, and Sandler 2017).

The MAF-ARF is an annual, person-level panel starting in 2000 that links each individual who can be assigned a Protected Identification Key (PIK) to the MAFID of the housing unit where that person appears in federal administrative records during the reference year.[31] It draws PIK-level address data from sources including IRS Forms 1040 and 1099, the Medicare Enrollment Database, Indian Health Service patient records, Selective Service registrations, and Housing and Urban Development's (HUD) Public and Indian Housing and Tenant Rental Assistance Certification System databases. Prior to 2012, all observed addresses per PIK-year are retained; from 2012 onward, one is selected at random. In 2019, the MAF-ARF covered approximately 305 million individuals—about 93 percent of the U.S. resident population (Sullivan and Genadek 2024).

Post 2011, the LEHD-ICF acquires its residential locations from internal versions of the LEHD Residence Candidates File (RCF).[32] The RCF methodology and sources broadly follow the MAF-ARF construction, but it models residence location probabilistically given address conflicts in the source files (Graham, Kutzbach, and Sandler 2017). The LEHD-ICF records only the most preferred location for each year in the RCF (Graham et al. 2022). We show in Appendix A.2 that this choice is associated with higher tract-level location matches with the ACS than for the MAF-ARF.

We construct PIKs' residential location histories by combining the MAF-ARF and LEHD-ICF. To do so, we first link PIKs across the 2000-2019 MAF-ARF and, separately, the 2012-2019 LEHD-ICF. We randomly retain one MAFID for PIKs with multiple MAFIDs in a given pre-2012 MAF-ARF year. We then smooth PIKs' MAFID histories separately in the LEHD-ICF and MAF-ARF by first carrying the nearest available MAFID forward or backward to fill any missing MAFID-year records. When a PIK appears at the same MAFID within a ten-year window, we treat them as residing continuously at that

[31] A PIK is an anonymized identifier assigned by the Census Bureau's Person Identification Validation System, which matches name, date of birth, sex, and, when available, Social Security number to the SSA's master file of person records—the Numident (Wagner and Layne 2014).

[32] Before 2012, the LEHD-ICF acquires its residential location histories from the Composite Person Record (CPR) source, produced by the Center for Administrative Records Research and Administration. We are unable to access address identifiers (MAFIDs) from the CPR, however, and the CPR has been found to differ in important ways from the MAF-ARF (Graham, Kutzbach, and Sandler 2017). We therefore rely solely on the MAF-ARF for residential locations prior to 2012.

MAFID during the intervening years. We truncate PIKs' residential histories whenever the Numident records a person's death before their history's nominal end. Finally, we overlay the smoothed MAF-ARF and LEHD-ICF residential histories, taking the LEHD-ICF as authoritative in any conflict and conduct the same smoothing procedure as we did separately for the MAF-ARF and LEHD-ICF. These choices help ensure high observed levels of residential mobility are not caused by sampling error in the Census residential address infrastructure.

Property Characteristics. We obtain property-level characteristics from the following sources: the 2017 vintage of CoreLogic's Tax History and Multiple Listing Service (MLS) files, which both span years 2000-2016, though coverage is sparse in earlier years; the 2019 and 2020 Black Knight Property Assessment files for these same years; the internal household-level ACS for years 2005-2021; and the 2019 MAF-X. From these sources, we primarily retain information related to each unit's occupancy status (rental vs. owner-occupied).

The Census Bureau has added property-level tax assessment data aggregated from county and township assessment records by CoreLogic and Black Knight into its Integrated Research Environment (IRE). These property assessments have been merged with the MAF where possible (Brummet 2014; Clark et al. 2023). For properties merged to the MAF, we record occupancy status in assessed years, which we understand is determined by comparing property owners' mailing addresses to the assessed property addresses (Brummet et al. 2016). The Census Bureau has similarly integrated CoreLogic's MLS aggregation into the IRE. We record from the MLS data which properties were classified as rental units at listing time.

From the household-level 2005-2021 ACS, we record whether current residents reported owning or renting the surveyed unit. We also record from the ACS gross rents and other property characteristics like home size, which we use to compute tract-level hedonic rents, detailed in Appendix A.3. Finally, in addition to geographic identifiers used to map households' residential locations to 2010-delineated Tabulation Census Tracts, we record from the MAF-X whether each property belonged to a multi-unit building. We discuss in Appendix A.2 how we combine these data to assign annual rental unit indicators to MAFIDs.

Employment, Earnings, and Workplace Establishments. We obtain person-level employment and earnings outcomes, as well as workplace-establishment characteristics for select states and for years 2001-2019 from the S2021_R2022Q4 snapshot of the administrative LEHD database (Graham et al. 2022).[33] We draw person-level employment and earnings data from the Employment and Job History Files (LEHD-EHF and LEHD-JHF). These files contain quarterly earnings for each job covered by state unemployment insurance programs, covering roughly 95% of employment in the US.[34] Earnings and employment are recorded by state UI programs at the firm level. The Census Bureau additionally imputes workers' physical

[33]States must independently approve external Census projects' access to their LEHD employment and jobs data. Our project has access to the following states' data: Arizona, California, Colorado, Connecticut, Delaware, Indiana, Iowa, Kansas, Maine, Maryland, Nebraska, Nevada, New Jersey, New Mexico, New York, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, South Dakota, Tennessee, Texas, Utah, Virginia, Washington, Wisconsin, and Wyoming.

[34]Notable exclusions from state UI coverage include independent contractors, the unincorporated self-employed, and Federal government workers (Graham et al. 2022).

establishments for workers employed at multi-establishment firms. To do so, the Census Bureau employs a probabilistic model to assign workers to specific establishments based on the distance between the worker's residence and candidate workplaces as well as the relative size of each establishment within multi-establishment firms sourced from the Quarterly Census of Employment and Wages (QCEW) (Abowd et al. 2009). The model is trained on data from Minnesota, which reports establishment employment information for workers. The LEHD-JHF provides outcomes of ten probabilistic imputations; we assign multi-establishment firm workers to specific establishments based on the modal imputation, splitting ties randomly. We sum earnings over quarters and job spells to obtain person-level annual earnings among jobs covered by the LEHD. We assign each worker to the establishment where they earned the most in a given year. Lastly, we obtain establishments' 6-digit NAICS industry codes and census tract locations from the Employer Characteristics File (LEHD-ECF), which is sourced primarily from the QCEW.[35]

Non-Economic Person Characteristics. We obtain non-economic characteristics from the LEHD-ICF, the person-level 2005-2021 ACS, and the SSA Numident file. The LEHD-ICF provides demographics—race, ethnicity, gender, date and place of birth, and education—compiled from the SSA Numident, the 2000 and 2010 Decennial Censuses, and the ACS. Because many person-characteristics in the LEHD-ICF are missing from the underlying source files, the Census Bureau imputes these values with a classification and regression tree described in Graham et al. (2022). We flag every imputed characteristic and note in the paper whenever such values are used to build variables or samples given concerns about imputation error. For persons appearing in the ACS, we observe additional characteristics such as parental status and commute time and sometimes obtain values that fill gaps or conflict with the LEHD-ICF; in any conflict, ACS values take precedence. We harmonize coding across ACS years and the LEHD-ICF so all non-economic characteristics are consistently recorded.

Low-Income Housing and Rental Support. We collect information on low-income housing and rental support from three main sources: the US Department of Housing and Urban Development's (HUD) Picture of Subsidized Households (POSH) database; HUD's national Low-Income Housing Tax Credit (LIHTC) database; and hand-collected rent control ordinances.

We construct annual counts of units subsidized by HUD for every 2010-delineated census tract from 2009-2019. We use HUD's publicly available Picture of Subsidized Households (POSH), which includes annual tract-level counts of Housing Choice Vouchers, Public Housing, Project-Based Rental Assistance, and properties under the Section 236 Preservation Program, wherein property owners may refinance on favorable terms conditional on maintaining cost-based rents. For years 2009-2013, we sum program-level units within each tract; from 2014 onward we adopt HUD's tract-level totals. We crosswalk the pre-2012 files from 2000 to 2010 tract boundaries weighting tracts by rental-unit counts. We use cross-sectional counts as controls in our descriptive analyses and hedonic rent regressions; we avoid analyzing changes in subsidized housing units over time given significant year-to-year variability in the POSH counts.

[35] Establishments' tract locations are recorded using 2020 Census tract boundaries. We crosswalk these tracts to 2010 Census tract boundaries randomly using probability weights based on land overlap measures between the 2010 and the 2020 boundaries. We crosswalk establishments to their same 2010 tract boundaries each year.

We use HUD's national LIHTC database to construct annual tract-level counts of units that are likely subject to affordability requirements during years 2008-2019. We record each building's allocation year, total number of low-income units, 2010 tract identifier, flags for Qualified Census Tract (QCT) status and whether HUD was still monitoring the property in 2023. We deem a building's units subject to affordability requirements if (i) it was allocated within the past 15 years, (ii) it is HUD-monitored as of 2023, or (iii) it was allocated 16-30 years ago and lies outside a QCT—a proxy for extended-use compliance; pre-1990 projects are excluded unless still monitored.[36] We multiply HUD's reported low-income unit count by this affordability indicator and sum across buildings to obtain tract-level totals for each year in 2008-2019. As with the POSH, we use cross-sectional LIHTC counts as controls in our descriptive analyses and hedonic rent regressions.

We additionally construct annual, tract-level estimates of the share of low-income renter households living in rent-controlled units. First, we hand-collect details on over 600 rent-control ordinances covering all incorporated and unincorporated places in our analysis for years 2000-2019. We record each ordinance's years of operation, maximum allowable annual rent increase, and building-age and size eligibility rules. We then merge these ordinance details with the household-level 2005-2019 ACS microdata, which reports the relevant building attributes. We assign units as rent-controlled when their attributes satisfy the applicable ordinance criteria. We finally compute for each tract and year, the share of low-income renters in our sample occupying such units, smoothing the data over three-year bands.

A.2. Sample Construction

Identifying Rental Units. Our focus on renters is possible given our distinct ability to classify housing-unit tenure through linking the CoreLogic, Black Knight, ACS, and MAF-X databases. To this end, we build a housing tenure indicator panel in the following stages. (i) Using the 2005-2021 household-level ACS, we mark a property as rental or owner-occupied according to the interviewed household's survey response. (ii) We label a unit as rental if its CoreLogic MLS record shows a "for-rent" listing that year. When the ACS and MLS disagree, we treat the unit as owner-occupied in the survey year and rental in the following year, capturing cases in which an owner subsequently lists the home for rent. This ACS-MLS database is our "ground truth", which we use to validate subsequent tenure imputations. (iii) To expand coverage, we use the CoreLogic Tax History files to label units with absentee owners as rental and those with owner-occupiers as non-rental—an assumption borne out by the data: only 12.2 percent of absentee-owner properties in CoreLogic are reported as owner-occupied in the ACS-MLS data. (iv) We further classify units as owner-occupied when Black Knight assessment files identify them as such.[37] After interpolating the resulting housing tenure flags, first forward and then backward, to fill missing years within MAFIDs, we assign any still-unclassified MAFID-years as rental if the unit is marked as a multi-unit building in the MAF-X. This imputation falsely labels rental units as owner-occupied units in just 9.2 percent of observations in the ACS-MLS data.[38]

[36] We form these restrictions based on our reading of Khadduri et al. (2012).
[37] We were unable to distinguish absentee-ownership status from missing data in the Black Knight database.
[38] Update with corrected false positive estimates.

Annual Samples. We define annual samples for each year between 2010 and 2019. To do so, we merge PIKs' residential histories, employment and earnings, and non-economic characteristics from the aforementioned datasets. For each each year between 2010 and 2019, we define samples by implementing the following restrictions:

(a) Persons must be between 25 and 65 years old and not residing in a group quarter, as labeled by the MAF-X;
(b) Persons must be residing in one of the 100 largest CBSAs located within the LEHD states we can access;[39]
(c) Persons must be household heads, defined as PIKs with the highest earnings in the MAFID-year, breaking ties randomly;
(d) Household heads must have positive earnings for the year;[40]
(e) Household heads must be residing in a MAFID with an adjusted household income in the bottom tercile of their CBSA and decadal age-band's income distribution;[41]
(f) Household heads must never earn more than \$150,000 (2010 dollars) in a single year during our analysis period.[42]

We finally subset household heads based on their reported race (Black and non-Black) and whether their MAFID is a rental unit in the sample year. We record the outcome of this sample selection procedure in Table A1. The structural estimation uses data formed by stacking the 2010-2019 cross-sectional samples, whereas the descriptive and welfare analyses track each household head from its sample year forward.

Sample Validation. Following Sullivan and Genadek (2024), we validate our residential panel against the ACS. The leftmost columns of Table A2 show the share of sample households whose census tracts match those in the ACS. These tract-level match rates are remarkably high, especially considering households' high mobility rates and point-in-time survey variation. Match rates are higher from 2012 onward, likely because of the LEHD's model-driven approach to choosing among candidate residential locations starting in that year. The rightmost columns report households' one-year, tract-level move rates. Like Sullivan and Genadek (2024), we find considerable variation in these rates over time. As this variation likely reflects sampling error, we interpret cross-year move-rate differences cautiously and structure our analyses to circumvent its effects, as we document throughout the text.

[39] We include a CBSA only when at least 95% of its resident workers commute to jobs in states for which we have LEHD data, measured using work- and home-tract locations from the person-level ACS. This threshold ensures our LEHD-based employment counts yield accurate market-access measures for every CBSA in the analysis.

[40] We restrict to workers with positive earnings in a year given the innate difficulty of distinguishing between poor and rich non-workers, self-employed workers, and workers employed by firms not covered by the LEHD.

[41] We consider four equally-sized age-bands for each sample year containing household heads aged 25-34, 35-44, 45-54, and 55-65. We follow the Census Bureau's equivalence-scale guidelines to compute adjusted household income. We calculate adjusted household income for childless households as $\text{THI}/\sqrt{(\#\text{Adults})}$; for a single parent with one child as $\text{THI}/1.8^{.7}$; for single parents with multiple children as $\text{THI}/(1.8 + .5 \times (\#\text{Children} - 1))^{.7}$; and for all other households as $\text{THI}/(\#\text{Adults} + .5 \times \#\text{Children})^{.7}$, where THI is total household income. We approximate households with PIKs residing in the same MAFID and partition adults and children based on whether they are below 18 years of age. We additionally smooth adjusted household income for each household head over the sample year and the previous two years.

[42] We include this restriction given temporal fluctuations in earnings among very high earners in the LEHD.

Table A1. Constructing 2010 Samples

	Restriction	Number MAFIDs	Sample Description
	No restriction	99,980,000	
(a)	Between 25 and 65 and not group quarters	84,410,000	Tract aggregates sample
(b)	100 largest CBSAs	38,880,000	
(c)	Household Heads	38,880,000	
(d)	Earnings > $0	31,760,000 (3,750,000)	
(e)	Low-income	10,480,000 (1,870,000)	
(f)	Earnings never > $150,000 (2010 USD)	9,350,000 (1,750,000)	RMA measure sample
(g)	Rental unit	3,270,000 (864,000)	Structural analysis sample
(h.i)	2011-observed, urban, & low-college-share tracts	868,000 (351,000)	Baseline mobility sample
(h.ii)	2019-observed, urban, & low-college-share tracts	796,000 (324,000)	Regression analysis sample
(h.iii)	Gentrifying, urban, & low-college-share tracts	677,000 (269,000)	Welfare analysis sample

Notes: Table A1 reports the number of MAFIDs retained in the 2010 sample after applying each successive restriction. Parenthetical counts show the number of MAFIDs retained for households with Black household heads. Samples for other years are defined analogously. Samples (h.i), (h.ii), and (h.iii) are subsets of sample (g). The 2011- and 2019-observed restrictions require observing the household heads' residential census tract in both 2010 and 2011 or 2019, respectively. The urban restriction requires residence in a tract closer to the Metropolitan Division CBD than at least 50% of that division's population. The low-college-share tracts restriction requires residence in a tract in the bottom quartile of the CBSA's 2010 college-attainment distribution. The rightmost column names the samples formed at each cleaning stage.

A.3. Variable Construction

Neighborhood Sociodemographics. Our descriptive analyses rely on short-term variation in neighborhoods' sociodemographic characteristics. Ordinarily, researchers use pooled 5-year ACS aggregates to define US neighborhoods' sociodemographic characteristics and changes thereof. These 5-year ACS aggregates are noisily estimated, however, and exacerbate classical measurement error in descriptive analyses. We instead utilize our administrative data and combine potentially unbiased, but imprecise, 5-year ACS tract-level estimates with biased, but more precise, tract-level estimates derived from our residential panel defined after applying restriction (a) in Table A1.[43] We document below and detail comprehensively in French and Gandhi (2025) our novel approach to constructing these annual tract-level sociodemographic aggregates.

We assume the pooled five-year ACS aggregates are unbiased and use them to bias-correct tract aggregates derived from our residential panel to obtain precise and unbiased annual sociodemographic

[43] Bond et al. (2014) and Brummet and Reed (2021) show that children, recent movers, immigrants, low-income, unemployed, male, non-college-educated, and group-quarter residents are less likely to receive a PIK, thus biasing any tract aggregate derived from the residential panel.

TABLE A2. Sample Validation

	ACS Census Tract Match		**One-Year Move Rate**	
Year	Non-Black	Black	Non-Black	Black
2010	.791	.775	.181	.185
2011	.790	.772	.303	.308
2012	.864	.846	.220	.228
2013	.862	.848	.208	.221
2014	.863	.847	.234	.257
2015	.873	.860	.151	.162
2016	.880	.874	.227	.229
2017	.875	.866	.173	.182
2018	.878	.868	.173	.183
2019	.887	.867	-	-

Notes: The leftmost columns of Table A2 report the share of sample households whose census tracts match those in the ACS. The rightmost columns of Table A2 report households' one-year, tract-level move rates. The samples correspond to those defined after applying restriction (g) in Table A1, with the ACS Census Tract Match sample defined only for household heads we can merge to the ACS.

aggregates. Specifically, we assume that the logit-transformed pooled 5-year ACS sociodemographic neighborhood shares are unbiased measures of the true logit-transformed shares,[44]

$$Y^*_{\text{ACS}} = Y^*_{\text{True}} + \epsilon, \tag{A1}$$

and that the logit-transformed biased measures of the same neighborhood shares constructed from the residential panel are related to the true logit-transformed shares via a linear function of covariates,

$$Y^*_{\text{Panel}} = Y^*_{\text{True}} + \beta X, \tag{A2}$$

where $Y^* = \log\left(\frac{Y}{1-Y}\right)$, ϵ is a normally distributed error with mean zero and constant variance, Y^*_{True} is the true logit-transformed shares, and X is a vector of covariates including logit-transformed tract-level shares of adults with college degrees, adults who report as female, white, Black, another race, white but not Hispanic, Hispanic, foreign born, white with a college degree, white with a college degree but not Hispanic, with positive earnings in the LEHD, log-transformed tract-level measures of mean earnings, median earnings, adult population size, and tract-level housing supply elasticity and occupancy rate estimates from Baum-Snow and Han (2023), a quadratic in tracts' distance to their Metropolitan Division's CBD, a cubic in the calendar year, and CBSA-level fixed effects.

[44]We construct the pooled 5-year ACS measures internally in the Census IRE using person-level weights and cross-walking to 2010-delineated census tracts using block-to-block crosswalks and population weights.

Substituting A1 in to A2 we obtain

$$Y^*_{\text{Panel}} - Y^*_{\text{ACS}} = \beta X + \epsilon$$

which we estimate via OLS and use to predict $\widehat{Y^*} = Y^*_{\text{Panel}} - \hat{\beta}X$. We finally apply the inverse logit transformation to $\widehat{Y^*}$ to obtain our unbiased and precise annual tract-level measures of neighborhood sociodemographic shares. We obtain analogous unbiased and precise estimates for tract-level measures of mean earnings, median earnings, total adult population, and mean adult age substituting in *log*-transformations to equations A1 and A2. We obtain measures of total populations in each sociodemographic group referenced by multiplying our unbiased and precise estimates of tract-level shares with our estimates for total tract adult population.

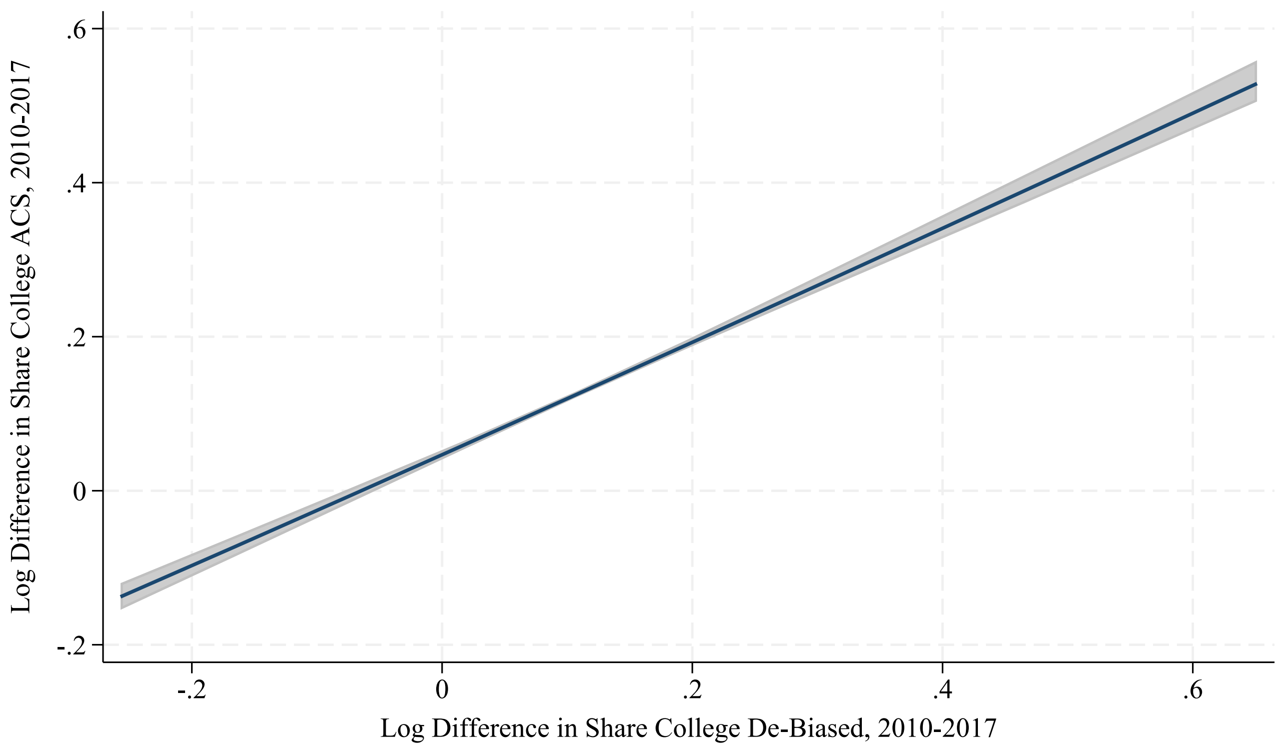
Panel A in Table A3 presents summary statistics comparing estimates of tracts' adult college-graduate shares computed using the ACS aggregates against these same shares computed using three estimates derived from our residential panel: estimates derived from all PIKs in our residential sample after applying restriction (a) in Table A1; estimates derived from these same PIKs but excluding those for whom the LEHD imputes educational attainment; and our de-biased tract estimates described above. Table A3 shows that using the full sample of PIKs in our residential panel to estimate the tract aggregates underestimates tracts' college-graduate shares, especially in later years. By contrast, we initially overestimate tracts' college-graduate shares when excluding from our residential panel PIKs whose educational attainment was imputed in the LEHD. These tract aggregates are significantly closer to their ACS counterparts than when we include PIKs with imputed educational attainment, however, as demonstrated by the markedly smaller Kullback-Leibler divergence measures. For this reason, we construct our de-biased college-graduate share tract aggregate estimates excluding PIKs with imputed educational attainment. As expected, our de-biased tract aggregates closely align with their ACS counterparts in both 2010 and 2017 according to our Kullback-Leibler divergence measures, the tract psuedo-medians, and the tract standard deviations.[45] Panel A additionally reports rounded pseudo-median counts of the number of PIKs used to compute each tract aggregate, rounded to the nearest 50 PIKs. The median de-biased tract aggregate is constructed using between 2.7 to 4.6 times more PIKs than its ACS counterpart, whose PIKs are further dispersed across five-year intervals.

Panel B in Table A3 presents similar summary statistics for log-differences in tract college-graduate shares between 2010 and 2017 for the pooled 5-year ACS aggregates and for our de-biased tract aggregates. While the pseudo-median tract aggregates are very similar across the two sets of estimates, their distribution—as indicated by the Kullback-Leibler divergence and standard deviation estimates—differ markedly. Smaller standard deviations among college-graduate tract shares derived from the de-biased panel implies meaningful sampling error in the ACS-based measures. While we believe our de-biased tract-level aggregates are a significant improvement over the pooled 5-year ACS estimates, they are not perfect; Figure A1 plots the results from a non-parameteric kernel regression of log-changes in neighborhood

[45]We present cross-section results for 2010 and 2017 so that the pooled 5-year ACS estimates centered at these years exclude years 2020 and 2021, since 5-year aggregates centered at 2018 and 2019 may be distorted by neighborhood re-sorting in response to COVID-19. In practice, we find very similar results when centering our pooled 5-year ACS estimates around 2018 and 2019. We nevertheless, estimate β in equation A2 excluding years 2018 and 2019, and conduct out-of-sample predictions for these same years.

college-graduate shares between 2010-2017 using the pooled ACS 5-year aggregates on the same measure constructed using the de-biased panel aggregates. That the kernel regression line is shallower than 45 degrees suggests some measurement error remains in the de-biased panel aggregates. We are happy with this... include the rent-ACS plots etc.

FIGURE A1. Assessing Measurement Error in De-Biased Panel Aggregates



Notes: Figure A1 plots results from the non-parameteric regression of log-changes in tracts' shares of adults holding college degrees between 2010-2017 using pooled ACS 5-year aggregates on the same measure constructed using the de-biased panel aggregates. We use epanechnikov kernals and select bandwidths via cross validation. 95% confidence interval bands are in gray.

Neighborhood Rents. We estimate two series of tract-level hedonic rents: one series tailored for our descriptive analyses to capture average changes in tract rents over time, and another series tailored for our structural and welfare analyses to capture changes in rents that incumbent renters actually pay, and therefore vary by the length of renters' neighborhood tenure. To construct the former series, we estimate tract-level hedonic rents by regressing real-2010 log gross rents from the household-level 2005-2021 ACS on lot size, indicators for the number of bedrooms, rooms, and the building type, as well as tract-by-year fixed effects. We then predict rents for a standardized property in each tract-year before smoothing these predictions over three-year bands, linearly interpolating them over missing values within census tracts, and winsorizing them at the first and ninety-ninth percentiles within each CBSA-year hedonic-rent distribution.[46]

[46]We predict hedonic rents using the approximation $\widehat{\text{Rent}}_{nt} \approx \exp\left(\ln\left(\widehat{\text{Rent}}_{nt}\right) + \frac{\text{Var}(\widehat{u}_{nt})}{2}\right)$, where $\widehat{u_{nt}}$ are residuals from the hedonic rent regressions.

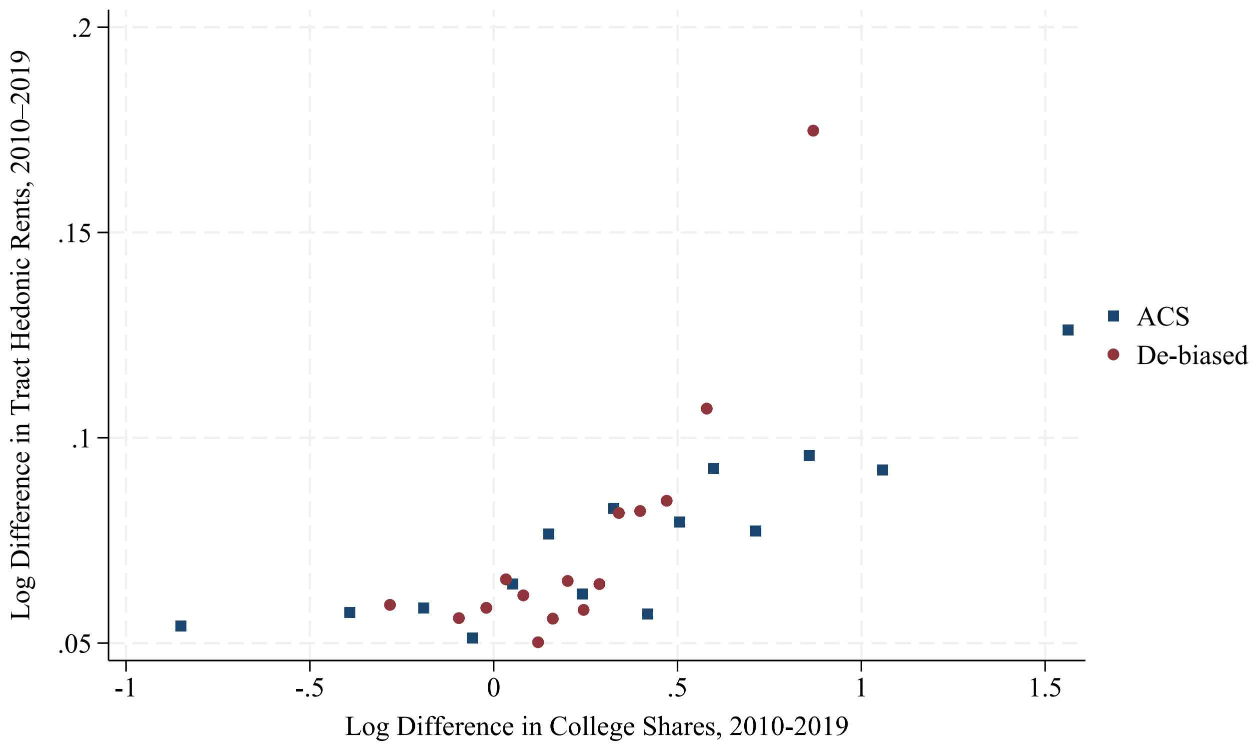
Table A3. Tract Aggregate Comparisons

Statistic	Measure	Estimate 2010	Estimate 2017
Panel A: Cross-Section Aggregates			
Kullback-Leibler divergence	College share all	.1428	0.2127
	College share not imputed	.0191	0.0031
	College share de-biased	.0039	0.0007
Tract pseudo-median	College share ACS	.2770	.3182
	College share all	.2670	.2583
	College share not imputed	.3038	.3125
	College share de-biased	.2744	.3127
Tract standard deviation	College share ACS	.2053	.2138
	College share all	.1444	.1338
	College share not imputed	.2033	.2134
	College share de-biased	.1968	.2137
Tract pseudo-median sample size	ACS	150	150
	All	2100	2200
	Not imputed	550	500
	De-biased	550	500
		2010-2017	
Panel B: Neighborhood Change Measures			
Kullback-Leibler divergence	Share college de-biased	.2714	
Tract pseudo-median	College share ACS	.1049	
	College share de-biased	.1039	
Tract standard deviation	College share ACS	.3530	
	College share de-biased	.1961	

Notes: Table A3 reports sample statistics for tract-level estimates of the share of residents aged between 25 and 65 who hold at least a Bachelors degree. These estimates are computed using (i) pooled 5-year ACS aggregates centered at 2010 and 2017 ("ACS"); (ii) all adults in our residential panel after applying restriction (a) of Table A1 ("All"); (iii) the "All" sample but excluding those for whom the LEHD imputes educational attainment ("Not Imputed"); and (iv) our de-biased tract-level estimates described in the appendix main text ("De-Biased"). Kullback-Leibler divergences are computed as $\sum_{x \in \mathcal{B}} P(x) \log(P(x)/Q(x))$, where $P()$ is the ACS distribution, $Q()$ the comparison distribution, and $\mathcal{B}$ the set of 20 bins we approximate these distributions with. Pseudo-median values equal the mean of the 11 tracts centered on the median tract—by college-educated share for the estimates and by PIK count for the sample sizes. Sample sizes correspond to the underlying number of PIKs used to compute each tract aggregate and are rounded to the nearest 50 per Census guidelines. Neighborhood change is measured in log differences.

We estimate the latter series tailored for our structural and welfare analyses similarly, but incorporate rental discounts stemming from (i) the differential propensity for sample households to reside in LIHTC

FIGURE A2. Validating De-Biased Panel Aggregates



Notes: Figure A2 shows the log change in tract-level hedonic rents as a function of the log change in the share of college-educated adults in sample households' 2010 residential tracts. The points plot conditional means estimated from a regression that includes CBSA fixed effects (Cattaneo et al. 2024). Tracts are weighted by the number of sample households residing in each tract in 2010. The sample comprises urban low-income renter households in 2010, defined after applying restriction (h.i) in Table A1. In the ACS series, we derive log differences in college shares only from the ACS tract estimates. In the de-biased series, we derive log differences in college shares from our de-biased tract estimates.

or otherwise subsidized housing units within tracts, and (ii) the share of sample households residing in rental units with binding rent control ordinances.[47] To do so, we additionally include in our hedonic rent regressions interactions between a "sample household" indicator and the tract counts of LIHTC units and, separately, other HUD-subsidized units detailed in A.1, as well as a four-way interaction between the sample indicator, an indicator denoting whether the unit has a binding rent control ordinance, the tract's share of low-income renters in rent-controlled units, and the household's tenure length. We estimate these hedonic regressions separately for each CBSA to accommodate heterogeneous impacts of LIHTC units, HUD subsidies, and rent control ordinances across metro areas.

We use these hedonic regressions to predict the gross rent sample households with a given tenure length would pay in each tract-year for a standardized property. The length of residential tenure only influences hedonic rents through our binding rent control measure, helping prevent reverse causality in which lower rents encourage households to remain longer in their units (Gallin et al. 2024). We similarly

[47]We define binding rent control ordinances as ordinances with maximum rent increases lower than their MSA's change in traditional, all transaction, Housing Price Index (HPI), retrieved from the Federal Housing Finance Agency given we do not observe counterfactual rent increases in the absence of rent control. When rent ordinances define maximum rent increases relative to the Consumer Price Index (CPI), we use MSA-wide CPIs to compute these maximum increases when possible, and regional CPIs when not, both obtained from the US Bureau of Labor Statistics.

smooth predictions over three-year bands, interpolate missing values within census tracts, and winsorize values at the fifth and ninety-fifth percentiles within each CBSA-year hedonic rent distribution (there is a higher share of outliers in this latter series than in the former series). For each tract-year, we finally aggregate predicted rents for households with tenure length less than or equal to five years ($\bar{\tau} = 1$) and for households with tenure length greater than five years ($\bar{\tau} = 2$).

Tract and Industry Sample Employment Counts. Our measures of residential and job market access depend on counts of low-income workers employed in each 6-digit NAICS industry in every 2010-delineated census tract. We construct these counts by recording workplace-tract locations and industries for low-income households defined aftera applying restriction (f) in Table A1 from the LEHD-ECF. We crosswalk industry definitions to their 2012-NAICS code using Census-provided concordance files and employment weights we construct from the administrative version of the Longitudinal Business Database.[48] Before conducting any analyses, we manually removed employment counts for industries in the LEHD-ECF whose establishments appeared to be improperly geocoded or assigned to headquarters that don't reflect employees' physical place of work.[49]

Appendix B. Microfoundations and Structural Estimation Details

B.1. Residential Market Access

Setup. In this section we derive our measure of residential market access introduced in Section 4.2 under the assumption that the labor market is segmented for each k-type worker.[50] Our derivation nests workers' choices across industries, which facilitates microfounding our job market access instruments described in Section 6. The workplace choice problem reported in Section 4.2 involves households choosing over both workplace, m, and industry, ι, conditional on living in neighborhood n:

$$I_{int} \equiv \max_{m,\iota} \frac{z_{im\iota t}}{d^k_{nmt}} \, w^k_{m\iota t}.$$

We assume z_{imt} is distributed iid Frechet over workplace locations and industries, $F(z_{im\iota t}) = \exp\left(-\left(z_{im\iota t}\right)^{-\epsilon^{kc}_t}\right)$ with shape parameter $\epsilon^{kc}_t > 1$ for each city c, period t, and for each type-k worker. The probability of a

[48] We follow establishments' changing industry classifications over time in the Longitudinal Business Database to infer the share of employment in each 2007- and 2017-delineated NAICS code that falls under each 2012-delineated NAICS code whenever industry definitions change. We ensure establishments' 2012-NAICS code assignments do not vary across years.

[49] These industries include: 112130—Dual-purpose cattle ranching and farming; 237310—Highway, street, and bridge construction; 491110, 488111, 485113, 482111, 48212, 485111, 492110—Postal services, Air traffic control, Bus and other motor vehicle transit systems, Railroads, Mixed-mode transit systems, Couriers and express delivery services; 561330, 541612, 541214, 541120—Professional employer organizations, Human resources consulting services, Payroll services, Offices of notaries; 611110—Junior and elementary schools; 813910, 813110, 814110, 813219—Business associations, Religious organizations, Private households, Other grantmaking and giving services; 92—Public services.

[50] 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 when assuming labor markets are segmented. 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).

type-k worker living in tract n taking a job in location m is then given by,

$$\pi^k_{m|nt} = \frac{\sum_\iota (w^k_{m\iota t}/d^k_{nmt})^{\epsilon^{kc}_t}}{\sum_\iota \sum_{m'} (w^k_{m'\iota t}/d^k_{nm't})^{\epsilon^{kc}_t}},$$

$$= \frac{\sum_\iota (w^k_{m\iota t}/d^k_{nmt})^{\epsilon^{kc}_t}}{\mathrm{RMA}^k_{nt}}, \tag{A3}$$

where $\mathrm{RMA}^k_{nt} \equiv \sum_\iota \mathrm{RMA}^k_{n\iota t} \equiv \sum_\iota \sum_{m'} (w^k_{m'\iota t}/d^k_{nm't})^{\epsilon^{kc}_t}$.

Define labor supply to tract m in time t by $\ell^k_{mt} = \sum_\iota \left[\left(w^k_{m\iota t}\right)^{\epsilon^{kc}_t}\right] \mathrm{FMA}^k_{mt}$, where FMA^k_{mt}, represents the access firms in tract m have to k-type workers in period t. Equating labor supply of type-k workers to tract m in period t to workers' choice probabilities yields an expression for FMA in terms of RMA:[51]

$$\ell^k_{mt} = \sum_n \pi^k_{m|nt} \cdot \pi^k_{nt} \cdot N^{kc}_t$$

$$= \sum_n \frac{\sum_\iota (w^k_{m\iota t}/d^k_{nmt})^{\epsilon^{kc}_t}}{\mathrm{RMA}^k_{nt}} \cdot \pi^k_{nt} \cdot N^{kc}_t$$

$$= \sum_\iota \left[\left(w^k_{m\iota t}\right)^{\epsilon^{kc}_t}\right] \cdot N^{kc}_t \cdot \sum_n \frac{\left(\left(\pi^k_{nt}\right)^{1/\epsilon^{kc}_t} / d^k_{nmt}\right)^{\epsilon^{kc}_t}}{\mathrm{RMA}^k_{nt}}$$

$$\equiv \sum_\iota \left[\left(w^k_{m\iota t}\right)^{\epsilon^{kc}_t}\right] \cdot \mathrm{FMA}^k_{mt} \tag{A4}$$

The penultimate equality yields,

$$\mathrm{FMA}^k_{mt} = N^{kc}_t \sum_n \frac{\left(\left(\pi^k_{nt}\right)^{1/\epsilon^{kc}_t} / d^k_{nmt}\right)^{\epsilon^{kc}_t}}{\mathrm{RMA}^k_{nt}}.$$

Furthermore, dividing both sides of A4 by $\left(d^k_{nmt}\right)^{\epsilon^{kc}_t}$ and summing over m yields an expression for RMA^k_{nt} in terms of FMA^k_{mt}. We subsequently obtain the following system of equations for RMA and FMA:

$$\mathrm{FMA}^k_{mt} = N^{kc}_t \sum_n \frac{e^{-\kappa^k \epsilon^{kc}_t \tau^k_{nmt}} \pi^k_{nt}}{\mathrm{RMA}^k_{nt}}$$

$$\mathrm{RMA}^k_{nt} = \sum_m \frac{e^{-\kappa^k \epsilon^{kc}_t \tau^k_{nmt}} \ell^k_{mt}}{\mathrm{FMA}^k_{mt}}, \tag{A5}$$

where we have defined $d^k_{nmt} \equiv e^{\kappa^k \tau^k_{nmt}}$ with τ^k_{nmt} denoting the commute time between tracts n and m for a

[51] N^{kc}_t is the total mass of type-k workers in city c and period t, and π^k_{nt} denotes the share of these workers choosing neighborhood n in period t.

type-k worker in period t and κ^k governing the magnitude of commute costs for type-k workers. [52] We solve for the unique (to scale) fixed point in A5 to obtain values for RMA^k_{nt} in each neighborhood, sample, and year (Tsivanidis 2022). We form measures of $N^{kc}_t \cdot \pi^k_{nt}$ and ℓ^k_{mt} directly from our samples defined after applying restriction (f) in Table A1. We describe how we obtain estimates of commute times (τ^k_{nmt}) and commuting elasticities ($\eta^{kc}_t \equiv \kappa^k \cdot \epsilon^{kc}_t$) below.

Commute Times. Period and k-type specific commute times relating all tract pairs within a CBSA (τ^k_{nmt}) play a crucial role in estimating RMA^k_{nt}. We obtain measures of τ^k_{nmt} from the ACS microdata for each n, m, t, and k combination. To do so, we match our samples to the ACS microdata and record their home census tract, n, their workplace census tract, m, and their reported commute time in minutes. We then compute the median reported commute time for each tract-pair with positive commute flows in each CBSA, separately for each sample and each five-year interval centered at every year between 2007 and 2017.[53]

Since there are far more potential tract-pair commuting routes within CBSAs than ACS survey respondents we can match to each sample, we are missing many observed period and k-type commute times. We therefore construct an empirical forecasting model to predict missing commute times between all neighborhood pairs within each CBSA for each sample and every year between 2001 and 2019. We follow Baum-Snow, Hartley, and Lee (2019) and predict these commute times using a CBSA fixed effect, the log-distance between the tract-pair centroids, and the log-distances between the home- and workplace-tract centroids and their respective metropolitan division CBDs. We estimate the following forecasting model separately for each sample and each year,

$$\ln\left(\tau^k_{nmt}\right) = \alpha^k_{d,t}\ \ln(\text{Dist})_{nm} + \alpha^k_{h,t}\ \ln(\text{Home CBD Dist})_n + \alpha^k_{w,t}\ \ln(\text{Work CBD Dist})_m + \alpha^{kc}_t + u^k_{nmt}, \tag{A6}$$

and construct $\widehat{\tau^k_{nmt}}$ using the approximation $\widehat{\tau^k_{nmt}} \approx \exp\left(\widehat{\ln\left(\tau^k_{nmt}\right)} + \frac{\widehat{\text{Var}(u^k_{nmt})}}{2}\right)$.

Table A4 shows that commute time responds less to additional travel distance for low-income renters than for the overall low-income population, with the elasticity smallest for Black households. This muted relationship among Black commuters is consistent with their lower car-ownership rates conditional on education and income (Raphael and Rice 2002). Commute times associated with public transit use are presumably less correlated with physical distance given heterogeneous public transit quality and availability between home- and workplace-tract pairs.

Commute Elasticities. In addition to tract-pair commute times, we must obtain estimates of the commute elasticities $\eta^{kc}_t \equiv \kappa^k \cdot \epsilon^{kc}_t$, which govern how the value of employment opportunities to workers decays with commute time. As η^{kc}_t increases, the more sensitive RMA^k_{nt} is to the proximity of workplace locations relative to these locations' employment levels, ℓ^k_{mt} (and thus equilibrium wages). We estimate η^{kc}_t through a series of gravity equations derived from the workplace choice model. Taking the (natural) log of equation

[52] $1 - e^{-\kappa^k \tau^k_{nmt}}$ thus represents the portion of time that type-k workers in tract n spend commuting to tract m in period t.
[53] We avoid using reported commute times from 2020 onward given the increased prevalence of working from home.

TABLE A4. Forecasting Commute Times

	Low-Income Renters		All Low-Income	
	Black	Non-Black	Black	Non-Black
Commute Distance $\left(\alpha^k_{d,t}\right)$	0.3565	0.4211	0.3717	0.4601
	(0.00449)	(0.002158)	(0.003202)	(0.00117)
Home-CBD Distance $\left(\alpha^k_{h,t}\right)$	-0.06569	-0.06897	-0.06462	-0.06631
	(0.005334)	(0.002484)	(0.003776)	(0.001379)
Work-CBD Distance $\left(\alpha^k_{w,t}\right)$	-0.06424	-0.07636	-0.06464	-0.07613
	(0.003375)	(0.001826)	(0.002289)	(0.0009395)
Adjusted R-Squared	0.2796	0.3472	0.2877	0.3928
N	27,000	91,000	51,000	286,000

Notes: Table A4 records the pseudo-median of the coefficient estimates of all CBSA-level regressions in the year 2010 from equation A6, separately for each sample. It also reports the mean of the standard errors and the adjusted R-squared statistics from the regressions forming the pseudo-median estimates. Each CBSA×year×sample regression is weighted by the number of commuters observed in each tract-pair cell in the ACS microdata.

A3 yields,

$$\begin{aligned} \ln\left(\pi^k_{m|nt}\right) &= \ln\left(\mathrm{RMA}^k_{nt}\right) + \ln\left(\sum_{\iota}\left(w^k_{m\iota t}\right)^{\epsilon^{kc}_t}\right) - \kappa^k \epsilon^{kc}_t \tau^k_{nmt} \\ &= \alpha^k_{nt} + \rho^k_{mt} - \eta^{kc}_t \tau^k_{nmt}. \end{aligned} \tag{A7}$$

We estimate these gravity equations separately for each CBSA, k-type worker, and period using observed commute flows in the LEHD to approximate $\pi^k_{m|nt}$ for each of our samples. Appendix XX describes how we assign workplace location to each sample household every period.[54]

Table A5 reports estimates of η^{kc}_t for our four low-income samples. As our estimates must include data from at least three of our LEHD-states, we are unable to display the distribution of η^{kc}_t across our sample CBSAs. We instead report three pseudo-quantiles (25^{th}, 50^{th}, 75^{th}) for each sample in 2010 and 2019. The pseudo-percentiles correspond to the mean of eleven estimates of η^{kc}_t centered at the given percentile in their distribution across CBSAs in a given year and sample. The elasticities in Table A5 are within the range of existing estimates, though on the higher end (Ahlfeldt et al. 2015). Like Baum-Snow, Hartley, and Lee (2019), we find substantial heterogeneity in commute elasticities across CBSAs, as revealed by the range of the pseudo-percentiles. We additionally find that low-income renters have higher commute elasticities than the broader low-income population, a pattern that likely reflects their relatively lower frictions to relocating in response to changing job market access. Further, Black households exhibit lower commute elasticities than non-Black households.[55] This is consistent with

[54] we may not need such a sentence if we have the data appendix first.

[55] We find similar patterns across Black and non-Black households when using commute distance as our key variable in equation A7.

Table A5. Gravity Regressions

		Low-Income Renters		All Low-Income	
		Black	Non-Black	Black	Non-Black
Panel A: 2010					
Commute Time $\left(\eta^{kc}_{2010}\right)$	25^{th} percentile	0.06047	0.08794	0.05685	0.08533
		(0.00375)	(0.000350)	(0.00106)	(0.000184)
	50^{th} percentile	0.08081	0.1021	0.07747	0.09581
		(0.00393)	(0.000288)	(0.00276)	(0.000146)
	75^{th} percentile	0.09981	0.1127	0.09628	0.1058
		(0.00319)	(0.000194)	(0.000858)	(0.0000844)
	N	736,000	1,939,000	1,477,000	6,082,000
Panel B: 2019					
Commute Time $\left(\eta^{kc}_{2019}\right)$	25^{th} percentile	0.06204	0.08289	0.06082	0.07929
		(0.00282)	(0.000491)	(0.00148)	(0.000215)
	50^{th} percentile	0.07518	0.09597	0.07401	0.09208
		(0.00196)	(0.000240)	(0.00111)	(0.000121)
	75^{th} percentile	0.08890	0.1101	0.08893	0.1022
		(0.00159)	(0.000255)	(0.00132)	(0.000121)
	N	967,000	2,527,000	1,837,000	7,079,000

Notes: Table A5 reports the pseudo-percentiles of the distribution of the parameter clusters within each sample separately. Panel A reports these estimates for the year 2010 and Panel B for the year 2019. Means of the standard errors from the regressions forming each pseudo-percentile are in parentheses. We weight each gravity regression by the number of commuters in each origin census tract. We use Poisson pseudo-maximum likelihood to account for sparse tract-pair commute flows (Correia, Guimarães, and Zylkin 2020).

Job Market Access. To transparently relate our job market access instruments discussed in Section 6 to the recent advances in the quasi-experimental shift-share literature, we also disaggregate by industry and take linear approximations of Equation A5 to obtain Equation 16:

$$\text{JMA}^k_{n\iota t} = \sum_{m \in \mathcal{N}^c \setminus n} e^{-\eta^{ck} \tau^k_{n,m}} l^{\,k}_{m\iota t}, \tag{A8}$$

These linear approximations relate to earlier notions of market potential in international and regional trade theory, which conceptualize the demand for goods in a given region as the sum of demands in surrounding regions, weighted by bilateral transportation costs (Harris 1954; Hanson 2005). It is common in the regional economics literature to take linear approximations of structural measures of market access (e.g., Donaldson and Hornbeck (2016) and Herzog (2021)).

B.2. Neighborhood Flow Utility

In this section we derive our expression for neighborhood expected flow utilities. Conditional on expected neighborhood incomes, households solve for the their optimal level of housing consumption in each neighborhood. Because there is no saving in the model, households' budget constraints bind with equality. Optimal housing consumption in neighborhood n for a household with τ years of tenure under Cobb-Douglas preferences in equation 3 is therefore

$$H^{k*}_{\tau nt} = \frac{(1-\beta^k_c)\cdot \bar{I}^k_{nt}}{r_{\tau nt}} \tag{A9}$$

Substituting A9 and our amenity parameterizations into equation 2 and then taking the (natural) log yields equation 5,

$$\begin{aligned} u^k_n\left(s^c_{it}\right) = \tilde{\alpha}^k_n + \tilde{\alpha}^k_t + \tilde{\beta}^k_I \ln\left(\bar{I}^k_{nt}\right) + \tilde{\beta}^k_r \log\left(r_{\tau nt}\right) + \tilde{\beta}^k_A \ln\left(\frac{\text{Coll}_{nt}}{\text{Pop}_{nt}}\right) \\ + \tilde{\beta}^k_\tau \ln(\tau_{it}) + MC^k_t(n_t, n_{it-1}) + \tilde{\xi}^k_{nt} + \varepsilon_{int} \end{aligned}$$

where

$$\begin{aligned} \tilde{\alpha}^k_n &= \left[\alpha^k_n + \beta^k_C \ln\left(\beta^k_C\right) + \left(1-\beta^k_C\right)\ln\left(1-\beta^k_C\right)\right]\frac{1}{\sigma^k} \\ \tilde{\alpha}^k_t &= \alpha^k_t \,/\, \sigma^k \\ \tilde{\xi}^k_{nt} &= \bar{\xi}^k_{nt} \,/\, \sigma^k \\ \tilde{\beta}^k_I &= 1/\sigma^k \\ \tilde{\beta}^k_A &= \beta^k_A \,/\, \sigma^k \\ \tilde{\beta}^k_r &= (\beta^k_C - 1) \,/\, \sigma^k \\ \tilde{\beta}^k_\tau &= \beta^k_\tau \,/\, \sigma^k \\ r_{\tau nt} &= r_{nt}\cdot h_n\left(\tau_{it}, \{\bar{\omega}^{kc}_t\}^t_{t-\tau}\right). \end{aligned}$$

B.3. First-Step Conditional Choice Probabilities

To see how equation 12 exactly captures the neighborhood choice problem of households in our dynamic model, start by considering our specification for a household's conditional value function defined in 7,

$$\begin{aligned} v^k_n\left(x_{it}, \bar{\omega}^{kc}_t\right) &\equiv \bar{u}^k_n\left(x_{it}, \bar{\omega}^{kc}_t\right) + \delta\mathbb{E}_t\left[\bar{V}^k\left(x_{it+1}, \bar{\omega}^{kc}_{t+1}\right)|n, x_{it}, \bar{\omega}^{kc}_t\right] \\ &= \tilde{\alpha}^k_n + \tilde{\alpha}^c_t + \tilde{\beta}^k_I \ln\left(\bar{I}^k_{nt}\right) + \tilde{\beta}^k_r \log\left(r_{nt}\right) + \tilde{\beta}^k_A \ln\left(\frac{\text{Coll}_{nt}}{\text{Pop}_{nt}}\right) + MC^k_t(n_t, n_{it-1}) + \tilde{\xi}^k_{nt} \\ &\quad + \sum_{\bar{\tau}=1,2}\Bigg(\delta\mathbb{E}_t\left[\bar{V}^k\left(x_{it+1}(n,\bar{\tau}), \bar{\omega}^{kc}_{t+1}\right)|n, x(n_{t-1},\bar{\tau}_{t-1}), \bar{\omega}^{kc}_t\right] \\ &\quad + \tilde{\beta}^k_\tau \ln(\bar{\tau}) + \tilde{\beta}^k_r \ln\left(h_n\left(\bar{\tau}, \{\bar{\omega}^{kc}_t\}^t_{t-\bar{\tau}}\right)\right)\Bigg)\cdot f^x(x(n_t,\bar{\tau})|n_t, x(n_{t-1},\bar{\tau}_{t-1})). \end{aligned} \tag{A10}$$

Let $\tilde{u}^k_{nt} \equiv \tilde{\alpha}^k_n + \tilde{\alpha}^c_t + \tilde{\beta}^k_I \ln\left(\bar{I}^k_{nt}\right) + \tilde{\beta}^k_r \log\left(r_{nt}\right) + \tilde{\beta}^k_A \ln\left(\frac{\text{Coll}_{nt}}{\text{Pop}_{nt}}\right) + \tilde{\xi}^k_{nt}$. Then, we can define the components of equation 12 as,

$$\begin{aligned}
\gamma^k_{nt\bar{\tau}=1} &\equiv \tilde{u}^k_{nt} + \delta\left[\mathbb{E}_t\left[\bar{V}^k\left(x_{it+1}(n,\bar{\tau}=1),\bar{\omega}^{kc}_{t+1}\right)|\bar{\omega}^{kc}_t\right] + \tilde{\beta}^k_r \ln\left(h_n\left(\bar{\tau}=1,\{\bar{\omega}^{kc}_t\}\right)\right)\right]\cdot\left(1-g^k_{nt}\right) \\
&\quad + \delta\left[\mathbb{E}_t\left[\bar{V}^k\left(x_{it+1}(n,\bar{\tau}=2),\bar{\omega}^{kc}_{t+1}\right)|\bar{\omega}^{kc}_t\right] + \tilde{\beta}^k_r \ln\left(h_n\left(\bar{\tau}=2,\{\bar{\omega}^{kc}_t\}\right)\right) + \tilde{\beta}^k_\tau \ln(2)\right]\cdot g^k_{nt}, \\
\gamma^k_{nt\bar{\tau}=2} &\equiv \tilde{u}^k_{nt} + \delta\left[\mathbb{E}_t\left[\bar{V}^k\left(x_{it+1}(n,\bar{\tau}=2),\bar{\omega}^{kc}_{t+1}\right)|\bar{\omega}^{kc}_t\right] + \tilde{\beta}^k_r \ln\left(h_n\left(\bar{\tau}=2,\{\bar{\omega}^{kc}_t\}\right)\right) + \tilde{\beta}^k_\tau \ln(2)\right] \\
\mu^k_{nt} &\equiv \tilde{u}^k_{nt} + \delta\left[\mathbb{E}_t\left[\bar{V}^k\left(x_{it+1}(n,\bar{\tau}=1),\bar{\omega}^{kc}_{t+1}\right)|\bar{\omega}^{kc}_t\right] + \tilde{\beta}^k_r \ln\left(h_n\left(\bar{\tau}=1,\{\bar{\omega}^{kc}_t\}\right)\right)\right] + MC^{kc} \\
&\quad + \left(MC^{kc}_{OO} - MC^{kc}\right)\cdot\mathbb{1}\left\{n = OO^c\right\} \\
\lambda^{kc} &\equiv MC^{kc}_{OO} - MC^{kc}
\end{aligned}$$

Having defined $\gamma^k_{nt\bar{\tau}=1}$, $\gamma^k_{nt\bar{\tau}=2}$, μ^k_{nt}, and λ^{kc} in this way, we can rewrite equation A10 as their linear combination plus our variable moving costs specification for movers within cities' urban cores:

$$\begin{aligned}
v^k_n\left(x_{it},\bar{\omega}^{kc}_t\right) = {}& \gamma^k_{nt\bar{\tau}=1}\cdot\mathbb{1}\{\bar{\tau}=1, n=\underline{n}\} + \gamma^k_{nt\bar{\tau}=2}\cdot\mathbb{1}\{\bar{\tau}=2, n=\underline{n}\} \\
&+ \mu^k_{nt}\cdot\mathbb{1}\{n\neq\underline{n}\} + \lambda^k\cdot\mathbb{1}\{\underline{n}=OO^c, n\neq\underline{n}\} \\
&+ \left(\boldsymbol{\beta}^{k\prime}_d\,\boldsymbol{d}(n,\underline{n}) + \boldsymbol{\beta}^{k\prime}_s\,\boldsymbol{s}(n,\underline{n})\right)\cdot\mathbb{1}\{n\neq OO^c, \underline{n}\neq OO^c\}
\end{aligned}$$

The right-hand-side of this equation matches the right-hand-side of the Poisson regression in 12. The frequency of moves to neighborhoods within each city's urban core identifies μ^k_{nt}, while the share of households of each aggregate tenure level staying in their home neighborhood exactly identifies $\gamma^k_{nt\bar{\tau}=1}$ and $\gamma^k_{nt\bar{\tau}=2}$. The share of households in each city's outside option moving to its urban core identifies λ^{kc}. Variation in the distance households move within their urban core, conditional on moving, identifies $\boldsymbol{\beta}^{k\prime}_d$. We discuss identifying households' variable moving costs in the main text.

The dependent variable in equation 12 is the number of sample household moves between tract pairs observed in the data. We detail how we count tract-pair moves and moves between tracts' outside options below. We weight equation 12 by the number of sample households from each aggregate tenure level in each year and origin-tract. We use the estimated Poisson regressions to predict the probability each k-type household moves between each neighborhood in their city. Table A6 reports regressions of the predicted conditional choice probabilities for movers on their empirical counterparts constructed from the raw data. As expected, the coefficients from this regression lie below one.[56] In the raw data, many neighborhood choices are never selected by low-income renters from a particular neighborhood, resulting in sparse empirical choice probabilities. The Poisson regression estimates, however, always assign some positive probability to each potential move. The larger coefficient for Non-Black households derives from their larger sample size and subsequent higher share of non-zero empirical choice probabilities.

[56] Additionally including households' home tracts in these regressions yields coefficient estimates very close to 1, but this is a poor description of the smoothing procedure given the highly bimodal distribution of move probabilities.

Table A6. Predicted Conditional Choice Probabilities

	Empirical Choice Probabilities	
	Black	Non-Black
Predicted Choice Probabilities	0.486	0.773
	(3.39e-06)	(1.96e-06)
Constant	0.0000912	0.0000418
	(8.09e-09)	(4.92e-09)
PIK-years	9,472,000	25,470,000

Notes: Table A6 reports regressions of sample households' empirical neighborhood choice probabilities on their counterparts predicted from equation 12. We restrict the sample to empirical choice probabilities involving moves from households' origin tracts to another tract within their CBSA's urban core or to their outside option. We weight observations in equation 12 by the number of sample households in the corresponding origin tract, year, and aggregate tenure level. We weight each outside option-year observation equal to the maximum weight of any other observation in the same CBSA, year, and aggregate tenure level. The raw count data is smoothed over three-year bands, as described in the main Appendix B.3 text. We report the total number of PIK-years in the raw data used to construct all empirical choice probabilities.

Constructing Migration Counts. We now describe how we form counts of moves between census tracts within CBSAs' urban cores and between CBSAs' urban cores and their outside options. These counts form our dependent variable in equation 12, and thus the basis of our structural estimation. We define a CBSA's urban core as the set of census tracts containing the 80% of the CBSA's population that lives closest to its Metropolitan Divisions' CBDs.[57] We define the outside option separately for each CBSA. For a given CBSA, we consider all sample households residing outside that CBSA's urban core in period t—but within the CBSA—as in that CBSA's outside option during the period (the CBSA's "outer band"). We also consider all sample households who move into a CBSA in period $t+1$—into or outside its urban core—as in that CBSA's outside option during period t.

For each year, sample, and CBSA, we count residential moves between tracts within the urban core, between the urban core and its outside option, and within the outside option. To do so, we record each sample household's tract of residence in period t (origin) and period $t+1$ (destination), and classify their moves into one of 12 categories outlined in Table A7. After assigning each sample household to a category, we define their aggregate tenure levels as $\bar{\tau} = 1$ if the household lived in a different tract or outside option in $t-5$ and $\bar{\tau} = 2$ if it remained in the same tract or outside option in $t-5$. We next sum over moves within each origin-destination location (as defined in Table A7), aggregate tenure level, and year.

Because our structural analysis relies on changes in households' location choices over time, we adjust these raw move counts to remove year-to-year noise in baseline move rates introduced by imperfections in

[57] Some large CBSAs include multiple Metropolitan Divisions. We define our CBSA-level urban cores as the tracts closest to Metropolitan Division CBDs to exclude rural areas lying between Metropolitan Divisions within a CBSA.

Table A7. Classification of Residential Moves

ID	Origin	Destination	Recorded Moves
1	Urban core	Urban core of same CBSA	Tract $\rightarrow$ Tract
2	Urban core	Outer band of same CBSA	Tract $\rightarrow$ OO
3	Urban core	Outside any sample CBSA	Tract $\rightarrow$ OO
4	Urban core	Outer band of another sample CBSA	Tract $\rightarrow$ OO OO-Dest $\rightarrow$ OO-Dest
5	Urban core	Urban core of another sample CBSA	Tract $\rightarrow$ OO OO-Dest $\rightarrow$ Tract-Dest
6	Outer band	Urban core of same CBSA	OO $\rightarrow$ Tract
7	Outer band	Outer band of same CBSA	OO $\rightarrow$ OO
8	Outer band	Outside any sample CBSA	OO $\rightarrow$ OO
9	Outer band	Outer band of another sample CBSA	OO $\rightarrow$ OO OO-Dest $\rightarrow$ OO-Dest
10	Outer band	Urban core of another sample CBSA	OO $\rightarrow$ OO OO-Dest $\rightarrow$ Tract-Dest
11	Outside any sample CBSA	Outer band of a sample CBSA	OO-Dest $\rightarrow$ OO-Dest
12	Outside any sample CBSA	Urban core of a sample CBSA	OO-Dest $\rightarrow$ Tract-Dest

Notes: Each row in Table A7 row classifies a residential move for a sample household. Moves are defined by (i) the census tract of residence in year t ("origin") and (ii) the census tract in year t+1 ("destination"). "Outer bands" are the tracts in a CBSA that contain the 20% of a CBSA's population living farthest from its Metropolitan Divisions' CBDs. The right-most column records the type(s) of residential move(s) associated with the sample household's origin and destination locations. "OO" denotes an outside option tract, and the suffix "Dest" indicates that the tract or outside option belongs to the move's destination CBSA. We exclude a 13$^{\text{th}}$ category—moves that both start and end outside every sample CBSA—because such households are not in our sample during that year. Households that do not move (i.e., remain in their origin tract or in the same outer band) are counted in categories 1 and 7. Moves described by 4, 5, 9, and 10, involve locations in two different CBSA samples; we duplicate these observations and assign one record to each CBSA.

the MAF-ARF and LEHD-ICF documented in Table A2. For each sample, we compute the unconditional nationwide, one-year, tract-level move probability for sample households in every year and for the entire 2010-2019 panel. We then form an annual scaling factor as the ratio of a given year's move probability to the pooled move probability. We multiply the yearly counts of households that remain in their origin tract or outside option by this scaling factor. This smoothing keeps baseline move probabilities consistent across years without affecting choice probabilities conditional on moving. Finally, for the Poisson regressions, we further smooth the data by pooling counts within three-year windows centered on each year from 2011 to 2018 (i.e., counts for year t include moves from $t-1$, t, and $t+1$).

B.4. Deriving our Estimating Equation

Consider the set of residential choices detailed in Section 5.2.[58] Given these residential choices, we start our derivation of equation 11 with an application of the Hotz-Miller inversion, which amounts to differencing equation 8 across two neighborhood choices in period t, n and OO^c:

$$
\begin{aligned}
\ln\left(\frac{p^k_{xnt}}{p^k_{xOO^ct}}\right) &= v^k_{xnt} - v^k_{xOO^ct} \\
&= \bar{u}^k_{xnt} - \bar{u}^k_{xOO^ct} + \delta\left(\mathbb{E}\left[\bar{V}^k_{xnt}\right] - \mathbb{E}\left[\bar{V}^k_{xOO^ct}\right]\right),
\end{aligned} \tag{A11}
$$

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

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

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 10:

$$
\mathbb{E}\left[\bar{V}^k_{xnt}\right] = \sum_{x'} \bar{V}(x', \bar{\omega}^{kc}_{t+1}) f^x\left(x'|n, x_{nt}\right) + e^{\bar{V}}(n, x_{nt}, \bar{\omega}^{kc}_t, \bar{\omega}^{kc}_{t+1}), \tag{A12}
$$

with

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

These realized continuation values permit 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 household-level state variables in A12 to our expression for the difference in conditional choice probabilities in A11 yields

$$
\begin{aligned}
\ln\left(\frac{p^k_{xnt}}{p^k_{xOO^ct}}\right) &= \bar{u}^k_{xnt} - \bar{u}^k_{xOO^ct} \\
&\quad + \delta\left(\sum_{x'} \bar{V}(x', \bar{\omega}^{kc}_{t+1})\, f^x\left(x'|n, x_t\right) - \sum_{x'} \bar{V}(x', \bar{\omega}^{kc}_{t+1})\, f^x\left(x'|OO^c, x_t\right)\right) \\
&\quad + \delta \cdot \tilde{e}(x_t, \bar{\omega}^{kc}_t, \bar{\omega}^{kc}_{t+1})
\end{aligned}
$$

where $\tilde{e}(x_t, \bar{\omega}^{kc}_t, \bar{\omega}^{kc}_{t+1})$ is the difference between the expectational errors when residing in neighborhood n relative to OO^c in period t,

$$
\tilde{e}(x_t, \bar{\omega}^{kc}_t, \bar{\omega}^{kc}_{t+1}) \equiv e^{\bar{V}}(n, x_t, \bar{\omega}^{kc}_t, \bar{\omega}^{kc}_{t+1}) - e^{\bar{V}}(OO^c, x_t, \bar{\omega}^{kc}_t, \bar{\omega}^{kc}_{t+1}).
$$

[58] This derivation mostly follows procedures used in the ECCP literature, particularly Diamond, McQuade, and Qian (2018), Davis et al. (2021), and Almagro and Domínguez-Iino (2022).

Next, using equation 9 to substitute in for $\bar{V}(x', \bar{\omega}^{kc}_{t+1})$, we obtain,

$$\begin{aligned}
&\ln\left(\frac{p^k_{xnt}}{p^k_{xOO^ct}}\right) + \delta\left(\sum_{x'} \ln\left(p^k_{x'\bar{n}t}\right) f^x\left(x'|n, x_t\right) - \sum_{x'} \ln\left(p^k_{x'\bar{n}t}\right) f^x\left(x'|OO^c, x_t\right)\right) \\
&= \bar{u}^k_{xnt} - \bar{u}^k_{xOO^ct} + \delta\left(\sum_{x'} v^k_{\bar{n}}\left(x', \bar{\omega}^{kc}_t\right) f^x\left(x'|n, x_t, \bar{\omega}^{kc}_t\right) - \sum_{x'} v^k_{\bar{n}}\left(x', \bar{\omega}^{kc}_t\right) f^x\left(x'|OO^c, x_t, \bar{\omega}^{kc}_t\right)\right) \\
&+ \delta \cdot \tilde{e}(x_t, \bar{\omega}^{kc}_t, \bar{\omega}^{kc}_{t+1})
\end{aligned}$$

Recall that $\bar{n}$ 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,

$$\begin{aligned}
&\ln\left(\frac{p^k_{xnt}}{p^k_{xOO^ct}}\right) + \delta\left(\sum_{x'} \ln\left(p^k_{x'\bar{n}t}\right) f^x\left(x'|n, x_t\right) - \sum_{x'} \ln\left(p^k_{x'\bar{n}t}\right) f^x\left(x'|OO^c, x_t\right)\right) \\
&= \bar{u}^k_{xnt} - \bar{u}^k_{xOO^ct} + \delta\left(MC^k_t(\bar{n}, n) - MC^k_t(\bar{n}, OO^c) + \tilde{e}(x_t, \bar{\omega}^{kc}_t, \bar{\omega}^{kc}_{t+1})\right)
\end{aligned}$$

Applying assumption (b) and substituting in for the neighborhoods' flow utilities provides an equation linear in our model parameters,

$$\begin{aligned}
&\ln\left(\frac{p^k_{xnt}}{p^k_{xOO^ct}}\right) + \delta\left(\sum_{x'} \ln\left(p^k_{x'\bar{n}t}\right) f^x\left(x'|n, x_t\right) - \sum_{x'} \ln\left(p^k_{x'\bar{n}t}\right) f^x\left(x'|OO^c, x_t\right)\right) \\
&\quad = \bar{\tilde{\alpha}}^k_n + \tilde{\alpha}^k_t + \tilde{\beta}^k_w \ln(\bar{I}^k_{nt}) + \tilde{\beta}^k_r \ln(r_{\tau nt}) + \tilde{\beta}^k_A \ln\left(\frac{\text{Coll}_{nt}}{\text{Pop}_{nt}}\right) \\
&\quad + \tilde{\beta}^k_\tau \left(\sum_{x'} \ln(\tau_{xt}(x')) f^x\left(x'|n, x_{xt}\right) - \sum_{x'} \ln(\tau_{xt}(x')) f^x\left(x'|OO^c, x_t\right)\right) \\
&\quad + MC^k_t(n, \underline{n}) - MC^k_t(OO^c, \underline{n}) + \delta\left(MC^k_t(\bar{n}, n) - MC^k_t(\bar{n}, OO^c)\right) \\
&\quad + \delta \cdot \tilde{e}(x_t, \bar{\omega}^{kc}_t, \bar{\omega}^{kc}_{t+1}) + \tilde{\xi}^k_{nt},
\end{aligned}$$

where $\bar{\tilde{\alpha}}^k_n = \tilde{\alpha}^k_n - \alpha^{kc}$.[59] To condense notation, we write this equation as

$$Y^k_{x\underline{n}n\bar{n}t} = \bar{\tilde{\alpha}}^k_n + \tilde{\alpha}^k_t + \tilde{\beta}^k_w \ln(\bar{I}^k_{nt}) + \tilde{\beta}^k_r \ln(r_{\tau nt}) + \tilde{\beta}^k_A \ln\left(\frac{\text{Coll}_{nt}}{\text{Pop}_{nt}}\right) + \tilde{\beta}^k_\tau \tilde{\tau}_{xt} + \widetilde{MC}^{kc}_t + \nu^k_{x\underline{n}n\bar{n}t},$$

where

$$\begin{aligned}
Y^k_{x\underline{n}n\bar{n}t} &\equiv \ln\left(\frac{p^k_{xnt}}{p^k_{xOO^ct}}\right) + \delta\left(\sum_{x'} \ln\left(p^k_{x'\bar{n}t}\right) f^x\left(x'|n, x_t\right) - \sum_{x'} \ln\left(p^k_{x'\bar{n}t}\right) f^x\left(x'|OO^c, x_t\right)\right) \\
\tilde{\tau}_{xt} &\equiv \sum_{x'} \ln(\tau_{xt}(x')) f^x\left(x'|n, x_{xt}\right) - \sum_{x'} \ln(\tau_{xt}(x')) f^x\left(x'|OO^c, x_t\right) \\
\widetilde{MC}^{kc}_t &\equiv MC^k_t(n, \underline{n}) - MC^k_t(OO^c, \underline{n}) + \delta\left(MC^k_t(\bar{n}, n) - MC^k_t(\bar{n}, OO^c)\right)
\end{aligned}$$

[59] $\tilde{\beta}^k_r \ln(r_{\tau nt})$ is also formally given by $\tilde{\beta}^k_r\left(\ln(r_{nt}) + \sum_{x'} \ln\left(h_n\left(\tau_{xt}(x'), \{\bar{\omega}^{kc}_t\}^t_{t-\tau}\right)\right) f^x\left(x'|n, x_{xt}\right)\right)$, but we suppress this notation here.

$$\nu^k_{x\underline{n}n\bar{n}t} \equiv \delta \cdot \tilde{e}(x_t, \bar{\omega}^{kc}_t, \bar{\omega}^{kc}_{t+1}) + \tilde{\xi}^k_{nt}.$$

This is the same equation reported in 11. Its empirical analogue is given by

$$\hat{Y}^k_{x\underline{n}n\bar{n}t} = \bar{\tilde{\alpha}}^k_n + \tilde{\alpha}^k_t + \tilde{\beta}^k_w \ln(\bar{I}^k_{nt}) + \tilde{\beta}^k_r \ln(\hat{r}_{\tau nt}) + \tilde{\beta}^k_A \ln\left(\frac{\text{Coll}_{nt}}{\text{Pop}_{nt}}\right) + \tilde{\beta}^k_\tau \hat{\bar{\tau}}_{xt} + \widehat{\widetilde{MC^F}}^{kc}_t + \nu^k_{x\underline{n}n\bar{n}t},$$

where,

$$\begin{aligned}
\hat{Y}^k_{x\underline{n}n\bar{n}t} &\equiv \ln\left(\frac{\hat{p}^k_{xnt}}{\hat{p}^k_{xOO^ct}}\right) + \delta\left(\sum_{x'} \ln\left(\hat{p}^k_{x'\bar{n}t}\right) f^x\left(x'|n, x_t\right)\right. \\
&\qquad \left. - \sum_{x'} \ln\left(\hat{p}^k_{x'\bar{n}t}\right) f^x\left(x'|OO^c, x_t\right)\right) - \widehat{\widetilde{MC^V}}^{kc}_t \\
\ln(\hat{r}_{\tau nt}) &\equiv \ln(r_{nt}) + \sum_{x'} \ln\left(h_n\left(\bar{\tau}_{xt}(x'), \{\bar{\omega}^{kc}_t\}\right)\right) f^x\left(x'|n, x_{xt}\right) \\
\hat{\bar{\tau}}_{xt} &\equiv \sum_{x'} \ln(\bar{\tau}_{xt}(x')) f^x\left(x'|n, x_t\right) - \sum_{x'} \ln(\bar{\tau}_{xt}(x')) f^x\left(x'|OO^c, x_t\right) \\
\widehat{\widetilde{MC^\upsilon}}^{kc}_t &\equiv \widehat{MC^\upsilon}^k_t(n, \underline{n}) - \widehat{MC^\upsilon}^k_t(OO^c, \underline{n}) + \delta\left(\widehat{MC^\upsilon}^k_t(\bar{n}, n) - \widehat{MC^\upsilon}^k_t(\bar{n}, OO^c)\right) \\
\nu^k_{x\underline{n}n\bar{n}t} &\equiv \delta \cdot \tilde{e}(x_t, \bar{\omega}^{kc}_t, \bar{\omega}^{kc}_{t+1}) + \tilde{\xi}^k_{nt}.
\end{aligned}$$

and where $\hat{\cdot}$ represents estimates from the first step. $\widehat{\widetilde{MC^\upsilon}}^{kc}_t$ is the difference in either the fixed, $\upsilon = F$, or variable, $\upsilon = V$, portion of moving costs.

B.5. Residential Paths

This subsection details the complete set of residential paths we use to estimate our second- and third-step equations and how each type of residential path helps identify our structural parameters.

Complete Set of Residential Paths. Figure A3 plots each pair of residential paths used in the structural estimation. Each pair of residential paths involves two hypothetical type-k households, with one of these households always choosing the outside option in period t. The other household either moves to a different neighborhood within their urban core (paths 1(a), 1(b), 3(a) and 3(b)) or stays in their current neighborhood (paths 2(a) and 2(b)). Residential paths 3(a) and 3(b) both start from within the city's outside option. Residential paths 1(a), 2(a), and 3(a) are initiated by incumbent residents with aggregate tenure length $\bar{\tau} = 1$, while residential paths 1(b), 2(b), and 3(b) are initiated by incumbent residents with aggregate tenure length $\bar{\tau} = 2$. Households with aggregate tenure length $\bar{\tau} = 1$ who stay in their current neighborhood in period t transition to aggregate tenure length $\bar{\tau} = 2$ with probability g^k_{nt}. Every path ends in period $t + 1$ with both hypothetical households convening in the same neighborhood within the urban core.

Identification of Fixed Moving Costs and Preference for Tenure. The additional residential paths in Figure A3 beyond those described in Figure 6 help identify households' fixed moving costs and preferences for residential tenure. Table A8 records differences in fixed moving costs and residential tenure utility for each pair of residential paths. The table shows that differences in the likelihood households choose the residential path not passing through the outside option in 1(a)/1(b) relative to in 2(a)/2(b) or in 3(a)/3(b) separately identifies the fixed moving costs for each CBSA given some δ. Variation across neighborhoods in the probability a household with aggregate tenure $\bar{\tau} = 1$ staying in the same origin neighborhood transitions to $\bar{\tau} = 2$ in the following period—i.e. g_n^k—identifies $\tilde{\beta}_\tau^k$. Note that only paths 2(a)/2(b) or 3(a)/3(b) are necessary to identify households' fixed moving costs or preferences over residential tenure. We therefore run robustness checks in Section C that exclude residential paths 3(a) and 3(b). These paths identify preferences over residential tenure from households' propensities to transition to $\bar{\tau} = 2$ while in their outside option, a source of variation less attuned to the model framework.

Table A8. Fixed Moving Costs and Residential Tenure

Residential Path	$\widetilde{\widetilde{MC^F}}^{kc}$	$\tilde{\beta}_\tau^k \cdot \hat{\bar{\tau}}$
Path 1(a)	$(1+\delta)\left(MC^{kc} - MC_{OO}^{kc}\right)$	0
Path 1(b)	$(1+\delta)\left(MC^{kc} - MC_{OO}^{kc}\right)$	0
Path 2(a)	$-MC_{OO}^{kc} + \delta\left(MC^{kc} - MC_{OO}^{kc}\right)$	$\tilde{\beta}_\tau^k \cdot g_n^k \cdot \ln(2)$
Path 2(b)	$-MC_{OO}^{kc} + \delta\left(MC^{kc} - MC_{OO}^{kc}\right)$	$\tilde{\beta}_\tau^k \cdot \ln(2)$
Path 3(a)	$MC_{OO}^{kc} + \delta\left(MC^{kc} - MC_{OO}^{kc}\right)$	$-\tilde{\beta}_\tau^k \cdot g_{OO_c}^k \cdot \ln(2)$
Path 3(b)	$MC_{OO}^{kc} + \delta\left(MC^{kc} - MC_{OO}^{kc}\right)$	$-\tilde{\beta}_\tau^k \cdot \ln(2)$

Notes: Table A8 records differences in fixed moving costs and residential tenure utility for each pair of residential paths displayed in Figure A3 and discussed in text B.5.

Restrictions on Residential Paths. Including all residential path pairs in the second step of the structural estimation is computationally infeasible. For a city with $N^c + 1$ potential origin locations in period $t-1$, households can choose among N^c destination census tracts in period t, and $N^c - 1$ renewal actions in period $t+1$. With two individual tenure states, the number of possible residential paths within a single city is $2 \times (N^c + 1) \times N^c \times (N^c - 1)$. For the largest urban core in our sample, New York-Newark-Jersey City, which contains roughly 3,760 census tracts, the total number of potential residential paths rises to over 100 billion. To address this computational challenge and to help credibly identify households' preference parameters, we restrict the set of residential paths used in the second (and third?) step of estimation as follows:[60]

[60] Must make clear in first-step estimation that we use all tract pairs.

(a) *Restrict set of origin locations in period* $t-1$: We restrict the set of origin tracts within each city to include the city's outside option and the top ten percent of census tracts with the highest share of k-type low-income renters in the city's urban core. Besides easing the computational burden, this restriction helps identify β_τ^k. Since we construct g_n^k separately for each origin tract, restricting to origin tracts with the highest sample shares reduces measurement error in g_n^k. This is important given we estimate β_τ^k via OLS in a third-step conditional on our second-step preference estimates.[61]

(b) *Restrict set of destination locations in period* t: We drop residential paths from the second estimation step based on the following criteria for destination tracts:

   (i) destination tracts must be above the bottom percentile of the nationwide distribution of tract adult population in 2011;

   (ii) destination tracts must be below the top quartile of the nationwide tract-level adult college-share distribution in 2011;

   (iii) destination tracts must be within the urban core, defined by population weighted distance to metropolitan division CBDs in 2011;

   (iv) destination tracts must be outside the top and bottom percentiles of the distributions for all three instrumental variables and log-differenced endogenous neighborhood characteristics, where the log-differences are between years 2011 and 2018.

   While also easing computation burden, the restrictions prevent the local-average treatment effect that identifies the preference parameters from being driven by outliers or by tracts that households are exceedingly unlikely to choose.

(c) *Restrict set of renewal actions in period* $t+1$: We restrict renewal tracts in period $t+1$ to those most likely to reset households' individual neighborhood tenures. A tract qualifies as a renewal action if it:

   (i) lies inside the urban core, defined by population weighted distance to metropolitan division CBDs in 2011;

   (ii) is 10-15 miles from its destination tracts centroid;

   (iii) is within .2 and .6 2011-population weighted distance from its Metropolitan Division CBD.

   We determined these cutoffs by comparing the subsequent tenure lengths of sample households moving to candidate renewal tracts from destination tracts within the urban core to those arriving from outside the urban core (the outside option). These cutoffs minimize the difference in average future tenure lengths between these groups of households.

Weighting Residential Paths. The number of residential paths selected for estimation scales with the cube of the number of tracts in a CBSA. As a result, unweighted second- and third-step regressions would

[61] Describe this third-step if not described clearly earlier.

overweight neighborhoods in larger CBSAs. We therefore weight each residential path according to,

$$w^k_{\bar{\tau},\underline{n}n\bar{n}} = \frac{|I^k_{\bar{\tau},n,2011}|}{|\left\{(\underline{n},\bar{n}) \in \mathbb{N}_{\underline{n}n\bar{n}}\right\}|},$$

where $\underline{n}n\bar{n}$ denotes the origin, destination, and renewal tracts defining a residential path, where $|I^k_{\bar{\tau},n,2011}|$ is the total number of sample households in destination tract n in 2011 with aggregate tenure $\bar{\tau}$, and where $|\left\{(\underline{n},\bar{n}) \in \mathbb{N}_{\underline{n}n\bar{n}}\right\}|$ is the total number of residential paths among the restricted set of residential paths, $\mathbb{N}_{\underline{n}n\bar{n}}$, that include n as their destination neighborhood. Each destination tract in the second- and third- estimation steps are therefore weighted in proportion to the number of sample households in 2011. These weights also help ensure that our estimation is not driven by large relative changes to initially small move probabilities, which are more susceptible to sampling bias.

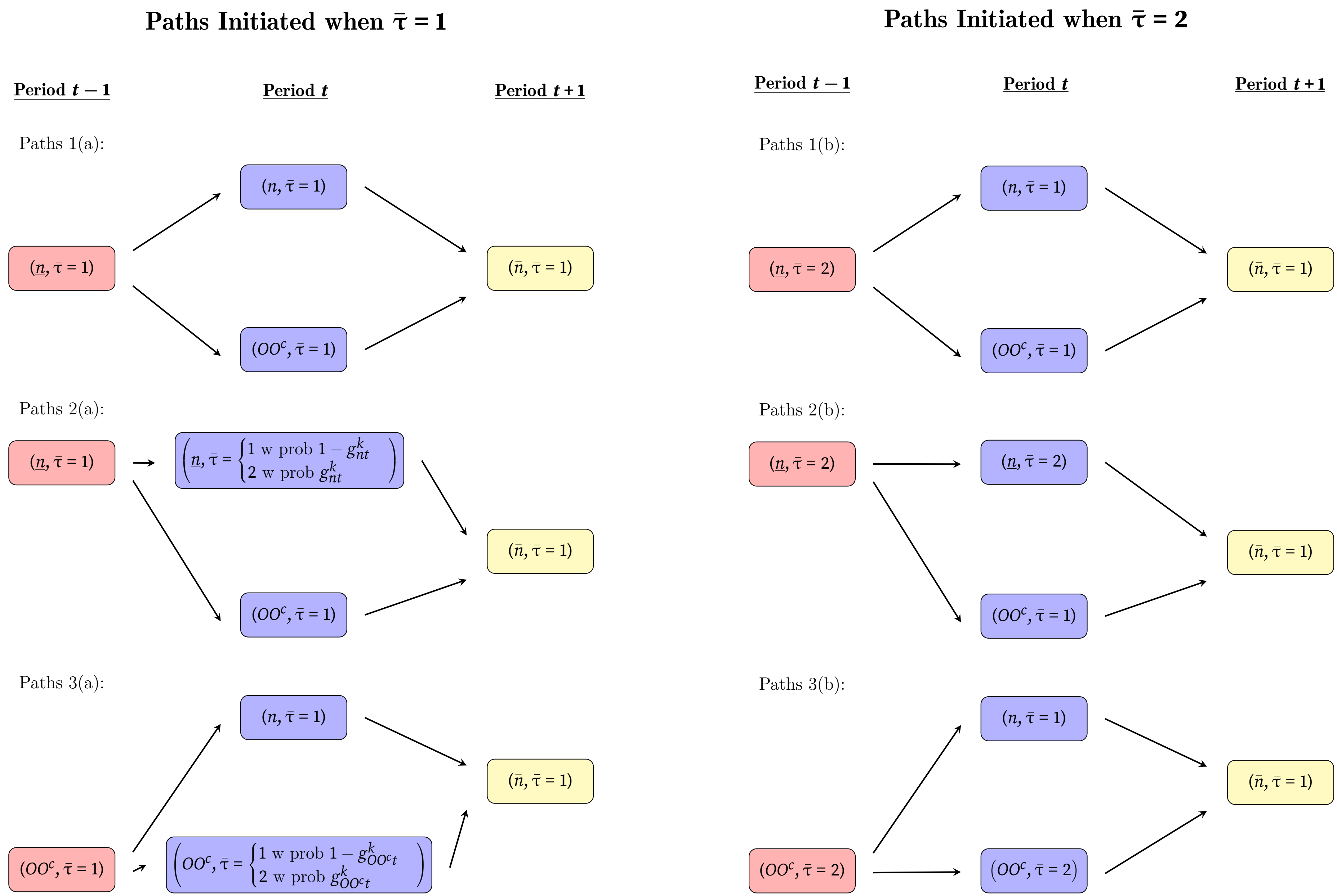


FIGURE A3. Residential Paths

Note: Figure A3 plots the residential paths we use to form estimating equations analogous to equation 11. Residential paths 1(a), 2(a), and 3(a) are initiated by incumbent residents in neighborhood $\underline{n}$ with aggregate tenure length 1; residential paths 1(b), 2(b), and 3(b) are initiated by incumbent residents with aggregate tenure length 2. We discuss each path's role in identification in the Appendix main text.

Appendix C. Identifying Equation ??

Formal discussion of identifying estimating equation coming soon.

Table A9. Complete Parameter Estimates

	OLS		IV	
	Black	Non-Black	Black	Non-Black
Endogenous variables				
Rent $\left(\tilde{\beta}_r^k\right)$	-0.0422 (0.0247)	.197 (.00707)	-2.633 (0.383)	-3.133 (0.280)
College share $\left(\tilde{\beta}_A^k\right)$	-0.187 (0.00669)	-.203 (0.00628)	0.485 (0.253)	1.293 (0.390)
Expected income $\left(\tilde{\beta}_w^k\right)$	-0.0434 (0.0235)	.0216 (.00860)	4.226 (0.684)	1.681 (0.684)
Continuous controls				
High tenure	-0.535 (1.363)	-1.095 (0.643)	-1.201 (2.062)	-0.952 (0.789)
Population-weighted CBD distance	-0.331 (0.053)	-0.357 (0.0556)	-0.622 (0.379)	0.882 (0.343)
Population-weighted CBD distance squared	0.415 (0.071)	0.406 (0.0629)	0.317 (0.358)	-0.851 (0.319)
Log college share	-0.0222 (0.0108)	-0.0147 (0.00711)	0.317 (0.0605)	0.427 (0.0505)
Log college share squared	0.00169 (0.00243)	-0.0307 (0.00181)	0.0247 (0.00717)	0.00653 (0.00408)
Log hedonic rents	-1.193 (0.122)	-5.713 (0.154)	-3.89 (0.49)	-11.12 (0.541)
Log hedonic rents squared	0.0964 (0.00933)	0.463 (0.0117)	0.237 (0.0298)	0.795 (0.0351)
White share	-0.691 (0.0294)	-0.502 (0.0227)	-0.608 (0.0975)	0.0447 (0.0394)
White share squared	0.928 (0.0254)	0.149 (0.0191)	0.474 (0.0786)	-0.508 (0.0811)
Black share	-0.232 (0.0245)	-0.582 (0.0238)	-0.957 (0.116)	-0.542 (0.0586)
Black share squared	0.205 (0.0192)	0.746 (0.0219)	0.62 (0.119)	0.633 (0.0302)
Urban development	-0.0621 (0.0104)	0.078 (0.00914)	0.0717 (0.0182)	0.302 (0.0191)
Urban development square	0.0557 (0.00762)	-7.94e-5 (0.00607)	0.0209 (0.0155)	-0.145 (0.016)
BHJ 2022 exposure controls				
Tradable employment total	2.09e-5 (2.41e-6)	3.79e-7 (8.61e-7)	1.49e-4 (3.93e-5)	2.09e-5 (1.82e-5)
Tradable employment share	-0.0829 (0.095)	1.236 (0.0819)	-0.281 (0.222)	-0.495 (0.408)
Tradable healthcare employment share	-0.204 (0.0788)	-0.696 (0.0768)	-0.871 (0.126)	-0.307 (0.101)
Tradable manufacturing employment share	0.302 (0.144)	-0.589 (0.0659)	0.144 (0.295)	0.297 (0.382)
Concentric population-weighted CBD-distance rings	✓	✓	✓	✓
Origin-tract fixed effects	✓	✓	✓	✓
SW F-stat (Rent)	—	—	93.69	42.66
SW F-stat (College share)	—	—	78.31	53.67
SW F-stat (Expected income)	—	—	83.26	45.57
Kleibergen-Paap F-stat	—	—	22.86	9.29
Residential paths (000s)	58,320	71,610	58,320	71,610

Notes: Table A9 extends the 2SLS estimates from Table 2 by including estimated coefficients for all continuous control variables and all corresponding OLS esimates. See notes to Table 2 for more details.

Table A10. Parameter Estimates and Sector Exposure

	(1)		(2)		(3)		(4)		(5)		(6)	
	Black	Non-Black	Black	Non-Black	Black	Non-Black	Black	Non-Black	Black	Non-Black	Black	Non-Black
Rent $\left(\tilde{\beta}_r^k\right)$	-2.42	-2.824	-2.125	-3.181	-2.079	-3.117	-2.423	-3.083	-2.633	-3.133	-2.592	-2.527
	(0.6)	(0.44)	(0.457)	(0.176)	(0.463)	(0.187)	(0.413)	(0.263)	(0.383)	(0.28)	(0.354)	(0.481)
College share $\left(\tilde{\beta}_A^k\right)$	1.101	1.451	0.138	1.562	0.133	1.457	0.344	1.247	0.485	1.293	0.363	0.0488
	(0.158)	(0.122)	(0.325)	(0.207)	(0.333)	(0.219)	(0.269)	(0.363)	(0.253)	(0.39)	(0.182)	(0.603)
Expected income $\left(\tilde{\beta}_w^k\right)$	0.606	0.543	5.374	0.987	5.304	1.265	4.776	1.683	4.226	1.681	3.739	4.46
	(0.33)	(0.268)	(0.946)	(0.246)	(0.958)	(0.298)	(0.758)	(0.646)	(0.684)	(0.684)	(0.532)	(1.114)
Origin-tract fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Baseline destination-tract controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
BHJ 2022 exposure controls												
Tradable employment total	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Tradable employment share			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Tradable healthcare employment share					✓	✓			✓	✓	✓	✓
Tradable manufacturing employment share							✓	✓	✓	✓	✓	✓
Each employment sector share											✓	✓
SW F-stat (Rent)	79.68	108.4	84.8	339.2	80.17	221.2	107.4	48.03	93.69	42.66	116.7	26.89
SW F-stat (College share)	76.21	115.7	58.83	260.8	56.5	204.5	78.92	57.26	78.31	53.67	131.9	31.65
SW F-stat (Expected income)	172.1	111.6	60.47	303.3	58.43	209.3	80.95	48.86	83.26	45.57	131.8	26.32
Kleibergen-Paap F-stat	17.94	35.8	18.65	48.61	17.87	39.32	24.27	10.38	22.86	9.29	31.89	5.741
Residential paths (000s)	58,320	71,610	58,320	71,610	58,320	71,610	58,320	71,610	58,320	71,610	58,320	71,610

Notes: Table A10 reports 2SLS estimates of the endogenous variables in Equation 14 for different market-access exposure controls following Borusyak, Hull, and Jaravel (2022). We discuss our choice of these controls in the main text of Appendix C. We describe the residential paths used in estimation and their weights in Appendix B.5. Standard errors, in parentheses, are clustered by origin tract. "SW F-stat" denotes the Sanderson–Windmeijer conditional F-statistic.

Table A11. Parameter Estimates Robustness Checks

	(1)		(2)		(3)		(4)		(5)	
	Black	Non-Black	Black	Non-Black	Black	Non-Black	Black	Non-Black	Black	Non-Black
Rent $\left(\tilde{\beta}_r^k\right)$	-1.846	-2.991	-2.548	-2.778	-2.675	-3.212	-3.977	-4.397	-2.271	-2.881
	(0.399)	(0.294)	(0.618)	(0.17)	(0.37)	(0.264)	(0.391)	(0.161)	(0.354)	(0.248)
College share $\left(\tilde{\beta}_A^k\right)$	0.444	1.342	0.391	1.181	0.501	1.306	-0.515	2.32	0.387	1.169
	(0.247)	(0.386)	(0.579)	(0.264)	(0.253)	(0.384)	(0.435)	(0.463)	(0.309)	(0.37)
Expected income $\left(\tilde{\beta}_w^k\right)$	4.201	1.638	5.008	1.298	4.214	1.715	7.612	1.493	4.072	1.839
	(0.672)	(0.719)	(3.32)	(0.926)	(0.679)	(0.691)	(1.236)	(1.137)	(0.723)	(0.607)
Origin-tract fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Baseline destination-tract controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
BHJ 2022 exposure controls										
Tradable employment total	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Tradable employment share	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Tradable healthcare employment share	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Tradable manufacturing employment share	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Robustness Checks										
Change in college-educated expected income	✓	✓								
Change in proximity to neighborhood college shares			✓	✓						
Exclude residential paths 3(a) and 3(b)					✓	✓				
Interact Bartik shift-share with proximity measure							✓	✓		
Change in racial shares									✓	✓
SW F-stat (Rent)	83.07	38.84	5.357	80.96	90.73	47.13	172.5	502.8	91.85	53.19
SW F-stat (College share)	75.29	51.71	4.709	37.78	76.65	56.87	53.25	41.68	60.1	63.15
SW F-stat (Expected income)	81.49	43.42	4.578	31.99	81.24	47.26	55.26	32.47	64.39	51.83
Kleibergen-Paap F-stat	22.12	9.109	1.466	6.436	22.46	9.546	11.79	5.66	17.61	11.22
Residential paths (000s)	58,320	71,610	58,320	71,610	58,030	71,270	58,320	71,610	58,320	71,610

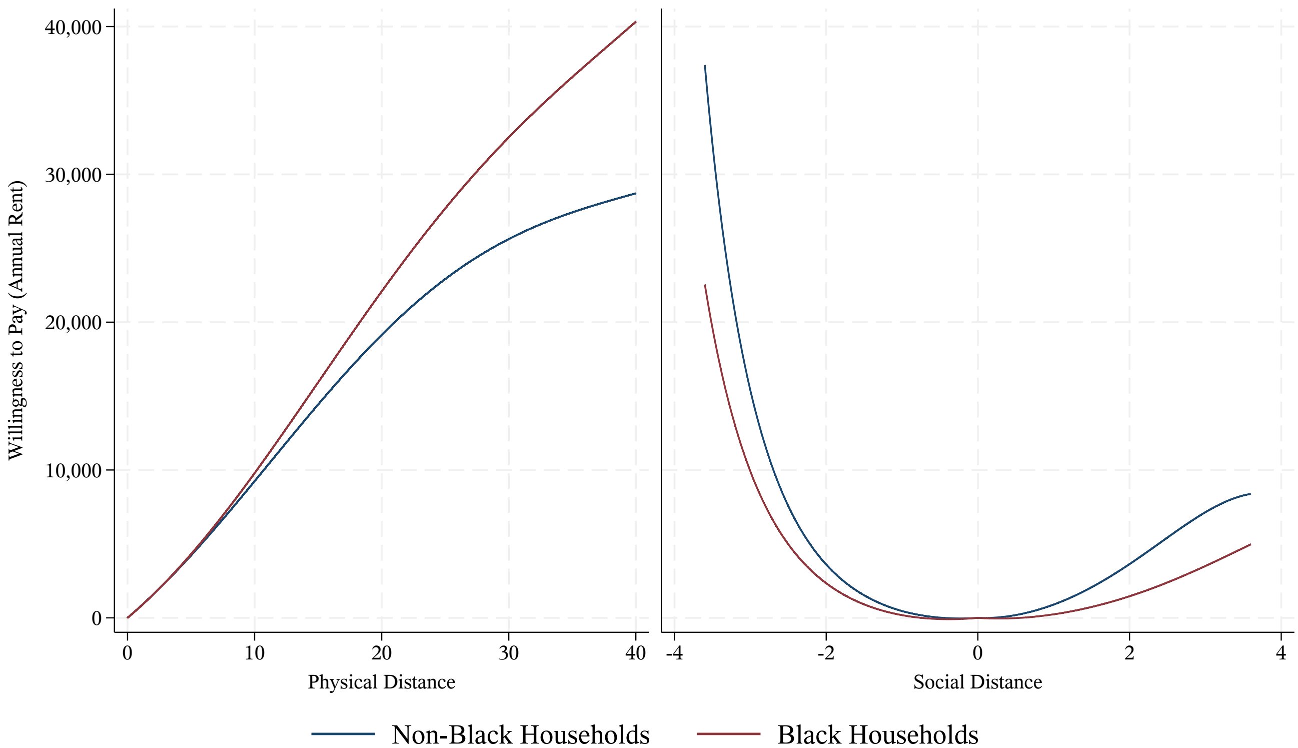
Notes: Table A10 reports 2SLS estimates of the endogenous variables in Equation 14 for four carefully selected robustness checks. We discuss these robustness checks in the main text of Appendix C. We describe the residential paths used in estimation and their weights in Appendix B.5. Standard errors, in parentheses, are clustered by origin tract. "SW F-stat" denotes the Sanderson–Windmeijer conditional F-statistic.

Table A12. Fixed Moving Cost Pseudo-Percentiles

	(1)						(2)					
	Black			Non-Black			Black			Non-Black		
Pseudo-percentile	25^{th}	50^{th}	75^{th}	25^{th}	50^{th}	75^{th}	25^{th}	50^{th}	75^{th}	25^{th}	50^{th}	75^{th}
Fixed tract moving cost $\left(MC^{kc}\right)$	-3.736	-3.319	-2.922	-3.587	-3.221	-2.948	-3.719	-3.308	-2.919	-3.526	-3.175	-2.907
Fixed CBSA moving cost $\left(MC^{kc}_{OO}\right)$	-3.784	-3.368	-2.938	-3.664	-3.348	-3.118	-3.707	-3.226	-2.853	-3.500	-3.147	-2.86
Exclude residential paths 3(a) and 3(b)							✓	✓	✓	✓	✓	✓
Residential paths (000s)		423,300			506,400			421,300			504,000	

Notes: Table A12 reports the pseudo-percentiles of the city-specific fixed moving cost parameters from Equation 13, computed conditional on the coefficient estimates in column (3) of Table 2 and column (3) of Table A11. Pseudo-percentiles are constructed as the mean of the five CBSA-specific parameter estimates whose ranks are centered on the indicated percentile of the CBSA distribution. We pool all cross-sections from 2010–2019 and use the same residential paths and weights as when estimating Equation 14. Column (2) drops residential paths that originate in a city's outside option (paths 3(a) and 3(b) in Figure A3). We could not compute standard errors for the fixed moving cost parameters due to computational constraints. We report the total number of residential paths used to estimate equation 13.

Figure A4. Variable Moving Cost Willingness to Pay Measures



Notes: Figure A4 translates the variable moving cost coefficients in Table 3 into households' willingness to pay, over and above any fixed moving costs, to avoid a move of a given distance within the same city. Physical distance is in miles; social distance is the log-difference between tracts' college-graduate shares. For reference, a social-distance value of 2 represents a move from a tract with a 9 percent college-graduate share (the sample average) to one with 66 percent.

Appendix D. Additional Descriptive Analyses

D.1. Baseline Residential Mobility

D.2. Gentrification and Residential Mobility

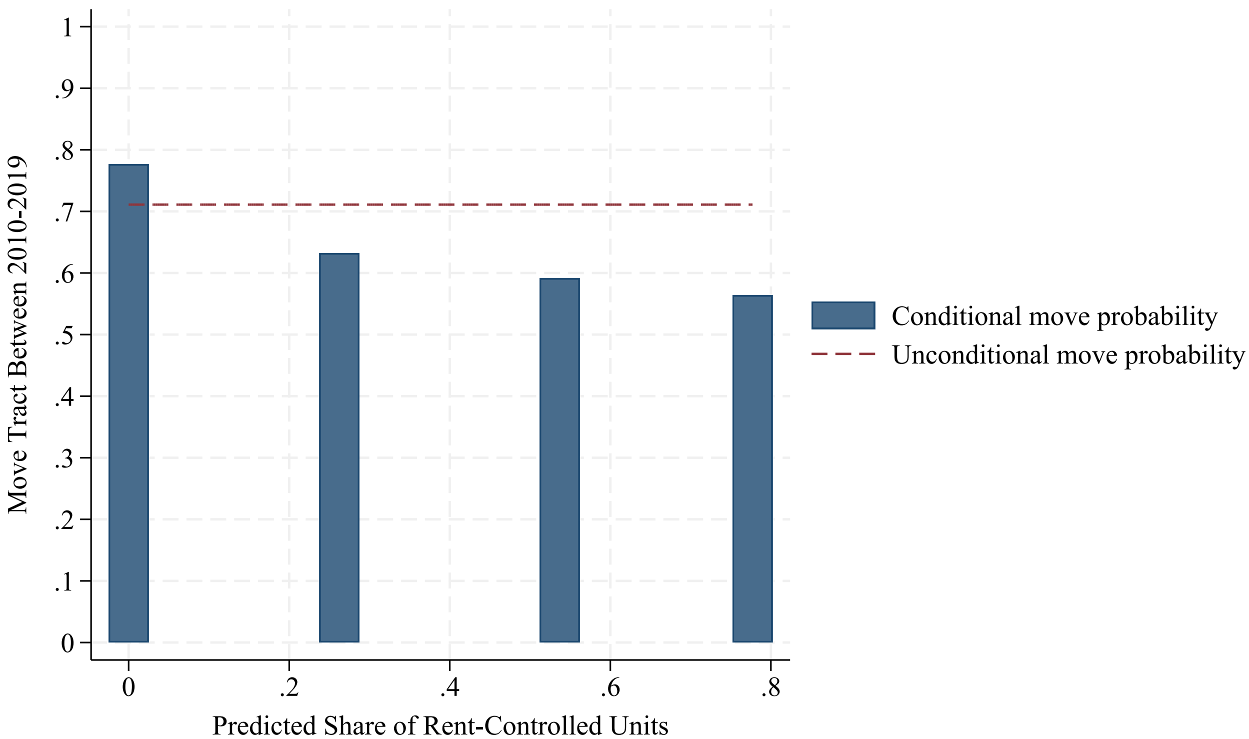
D.3. Gentrification and Experienced Neighborhood Characteristics

Table A13. Baseline Residential Mobility

	(1) Urban Low-Inc Renters		(2) Low-Inc Renters		(3) All Low-Inc HHs		(4) All HHs
	Non-Black	Black	Non-Black	Black	Non-Black	Black	
Panel A							
Move tract in 2010	0.1849	0.1869	0.1808	0.1844	0.1139	0.1457	0.1025
Move CBSA in 2010	0.03401	0.02230	0.04221	0.02715	0.02800	0.02251	0.02473
Move > 1 mile in 2010	0.1750	0.1770	0.1742	0.1770	0.1105	0.1400	0.09972
Move > 5 miles in 2010	0.1060	0.1014	0.1180	0.1087	0.07758	0.08832	0.07212
First-move distance	7.734	7.189	9.118	7.970	9.748	8.254	10.31
Standard deviation	(9.316)	(8.021)	(10.42)	(8.639)	(10.90)	(8.870)	(10.89)
Pseudo-median	[4.702]	[4.768]	[5.691]	[5.244]	[6.192]	[5.457]	[6.756]
Skew	⟨3.67⟩	⟨4.675⟩	⟨3.459⟩	⟨3.649⟩	⟨3.330⟩	⟨3.458⟩	⟨2.884⟩
Panel B							
Move tract between 2010-2019	0.6984	0.7298	0.7068	0.7269	0.5409	0.6307	0.5207
Distance from 2010 tract in 2019	8.687	7.854	10.07	8.726	10.88	9.107	11.77
Standard deviation	(10.50)	(8.920)	(11.23)	(9.522)	(11.69)	(9.771)	(12.00)
Pseudo-median	[5.299]	[5.158]	[6.429]	[5.724]	[7.135]	[6.061]	[7.956]
Skew	⟨3.601⟩	⟨3.901⟩	⟨2.894⟩	⟨3.179⟩	⟨2.769⟩	⟨3.088⟩	⟨2.495⟩
Panel C							
Number moves in 2010–2019	1.145	1.395	1.040	1.309	0.7246	1.084	0.6345
Standard deviation	(1.289)	(1.437)	(1.214)	(1.393)	(1.072)	(1.329)	(0.9839)
Pseudo-median	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Skew	⟨1.315⟩	⟨1.131⟩	⟨1.374⟩	⟨1.180⟩	⟨1.806⟩	⟨1.397⟩	⟨1.940⟩
N	517,000	351,000	1,994,000	799,000	5,866,000	1,594,000	24,830,000

Notes: Table A13 summarizes residential mobility for households in 2010. Panel A compares each household head's 2010 residential tract to their 2011 residential tract; Panel B compares their 2010 residential tract to their 2019 residential tract; Panel C reports the number of residential tract moves between 2010 and 2019. Columns (1), (2), (3), and (4) use the samples defined after applying restrictions (h.i), (g), (f), and (d) in Table A1, respectively. Household counts correspond to Panel A (counts for Panels B and C are similar). We compute residential mobility statistics only when the household head's census tract is observed in both years relevant to that panel (2010 and 2011 for Panel A; 2010 and 2019 for Panels B and C). The move-distance statistics are computed only for households we observe in the same CBSA in the end year of our comparisons. "First-move distance" and "Distance from 2010 tract in 2019" are measured in miles and are computed only for households who leave their origin census tract.

Figure A5. Baseline Residential Mobility and Rent Control



Notes: Figure A5 shows the probability incumbent renters move census tracts between 2010-2019 as a function of their origin tract's predicted share of rent-controlled units. The points plot conditional means estimated from a regression specification that includes the same baseline fixed effects and individual- and tract-level controls as in *list regression spec here*. The number of bins is chosen to minimize integrated mean squared error (IMSE) (Cattaneo et al. 2024). Construction of the predicted rent-controlled shares is described in Section A.1. The sample comprises all urban low-income renter households observed in both 2010 and 2019, defined after applying restriction (h.ii) in Table A1.

Table A14. Gentrification and Residential Mobility

	(1) Baseline		(2) Gentrifying Tracts		(3) Rent-Controlled Tracts		(4) Long-Term Renters		(5) Movers	
	Non-Black	Black	Non-Black	Black	Non-Black	Black	Non-Black	Black	Non-Black	Black
Panel A: Moved Tract OLS										
Coefficient estimate	0.011	0.017	0.021	0.029	0.007	0.011	0.003	0.023	—	—
Standard error	(0.005)	(0.005)	(0.006)	(0.007)	(0.006)	(0.005)	(0.007)	(0.008)	—	—
Baseline controls	✓	✓	✓	✓	✓	✓	✓	✓		
CBSA fixed effects	✓	✓	✓	✓	✓	✓	✓	✓		
Panel B: Moved Tract IV										
Coefficient estimate	0.015	0.027	0.020	0.033	0.019	0.017	0.009	0.027	—	—
Standard error	(0.007)	(0.007)	(0.009)	(0.009)	(0.007)	(0.006)	(0.009)	(0.010)	—	—
Kleibergen-Paap F-stat	[101.4]	[413.3]	[96.80]	[437.7]	[38.98]	[190.7]	[142.5]	[466.0]	—	—
Baseline controls	✓	✓	✓	✓	✓	✓	✓	✓		
CBSA fixed effects	✓	✓	✓	✓	✓	✓	✓	✓		
Panel C: Moved CBSA IV										
Coefficient estimate	0.013	0.006	0.015	0.004	0.027	0.011	0.000	-0.001	—	—
Standard error	(0.004)	(0.004)	(0.005)	(0.005)	(0.007)	(0.005)	(0.005)	(0.005)	—	—
Kleibergen-Paap F-stat	[101.4]	[413.3]	[96.8]	[437.7]	[38.98]	[190.7]	[142.5]	[466]	—	—
Baseline controls	✓	✓	✓	✓	✓	✓	✓	✓		
CBSA fixed effects	✓	✓	✓	✓	✓	✓	✓	✓		
N	472,000	324,000	389,000	255,000	260,000	219,000	201,000	127,000	—	—
Panel D: Distance from Origin Tract IV										
Coefficient estimate	0.078	0.081	0.100	0.101	0.088	0.057	0.049	0.055	0.099	0.043
Standard error	(0.015)	(0.015)	(0.019)	(0.020)	(0.020)	(0.017)	(0.019)	(0.021)	(0.014)	(0.015)
Kleibergen-Paap F-stat	[94.46]	[406.1]	[90.06]	[432.9]	[33.34]	[184.5]	[149.3]	[466.8]	[88.67]	[395.1]
Baseline controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
CBSA fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Panel E: Distance from CBD IV										
Coefficient estimate	0.012	0.035	0.033	0.057	0.052	0.027	-0.019	0.012	0.012	0.039
Standard error	(0.019)	(0.015)	(0.023)	(0.018)	(0.030)	(0.020)	(0.015)	(0.013)	(0.027)	(0.021)
Kleibergen-Paap F-stat	[94.46]	[406.1]	[90.06]	[432.9]	[33.34]	[184.5]	[149.3]	[466.8]	[88.67]	[395.1]
Baseline controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
CBSA fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
N	392,000	278,000	324,000	219,000	212,000	190,000	177,000	114,000	249,000	190,000

Notes: Table A14 reports coefficient estimates for our gentrification measure from regression equation **??**. In Panels A and B, the dependent variable is an indicator equal to one if the household resides in a different census tract in 2019 than in 2010. In Panel C, the dependent variable is an indicator equal to one if the household resides in a different CBSA in 2019 than in 2010. In Panel D, the dependent variable is one plus the log of the straight-line distance in miles from the centroid of the household's 2019 tract to the centroid of its 2010 origin tract. In Panel E, the dependent variable is one plus the log change in the household's distance in miles from its Metropolitan Division CBD to its tract centroid between 2010 and 2019. In Panels B-E, we instrument for gentrification using changes in the adult college-educated population, as described in Section **??**. All specifications include CBSA fixed effects and the full set of baseline controls listed in Table **??**. We define gentrifying tracts as those whose college share rose from 2010 to 2019; rent-controlled tracts as those covered by a rent-control ordinance in 2010; long-term renters as households residing in the same tract since at least 2005; and movers as households observed in a different census tract in 2019 than in 2010. Robust standard errors clustered at the origin census-tract level are in parentheses (Abadie et al. 2023). Kleibergen-Paap F-statistics are in square brackets. The sample comprises urban low-income renter households observed in both 2010 and 2019, defined after applying restriction (h.ii) in Table A1. Panels D and E further subset to households who are observed to reside in the same CBSA in 2019 as in 2010. All coeficient estimates and standard errors are rounded to three decimal places, while Kleibergen-Paap F-statistics are rounded to four significant figures.

Table A15. Gentrification and Residential Mobility Heterogeneity

	Non-Black			Black			Baseline Controls	CBSA FEs
	Coefficient	F-Stat	N	Coefficient	F-Stat	N		
Panel A: Alternative Gentrification Measures								
ACS 5-year Aggregates	0.005 (0.003)	23.57	472,000	0.008 (0.003)	206.8	324,000	✓	✓
White College Share	0.011 (0.005)	85.40	472,000	0.014 (0.004)	245.2	324,000	✓	✓
Panel B: Additional Sample Restrictions								
Inner City	0.011 (0.009)	64.27	293,000	0.018 (0.008)	221.5	209,000	✓	✓
No College Educated	0.014 (0.007)	95.98	425,000	0.030 (0.007)	432.1	293,000	✓	✓
Drop Outliers	0.009 (0.008)	1457	427,000	0.023 (0.009)	927.6	289,000	✓	✓
Below Median College Share	0.012 (0.007)	278.6	795,000	0.020 (0.007)	532.7	483,000	✓	✓
Inelastic Tracts	0.024 (0.010)	258.1	246,000	0.021 (0.009)	202.1	169,000	✓	✓
Elastic Tracts	0.008 (0.010)	43.22	225,000	0.032 (0.010)	217.3	155,000	✓	✓
Older Households	0.008 (0.011)	107.5	243,000	0.032 (0.009)	393.4	156,000	✓	✓
Panel C: Additional Controls								
Gentrification 2005-2009	0.017 (0.007)	101.6	472,000	0.028 (0.007)	415.4	324,000	✓	✓
Gentrification 2005-2009, Long-Term Renters	0.011 (0.009)	143.7	201,000	0.029 (0.010)	470.5	127,000	✓	✓
Log Change in Adult Population	0.012 (0.007)	92.84	472,000	0.028 (0.007)	399.4	324,000	✓	✓
Log Change in Tradable Market Access	0.015 (0.007)	100.7	472,000	0.027 (0.007)	412.7	324,000	✓	✓
Log Change in Racial Shares	0.015 (0.008)	90.55	472,000	0.023 (0.007)	363.5	324,000	✓	✓
Log Change in Hedonic Rents	0.002 (0.007)	96.63	472,000	0.018 (0.007)	399.1	324,000	✓	✓

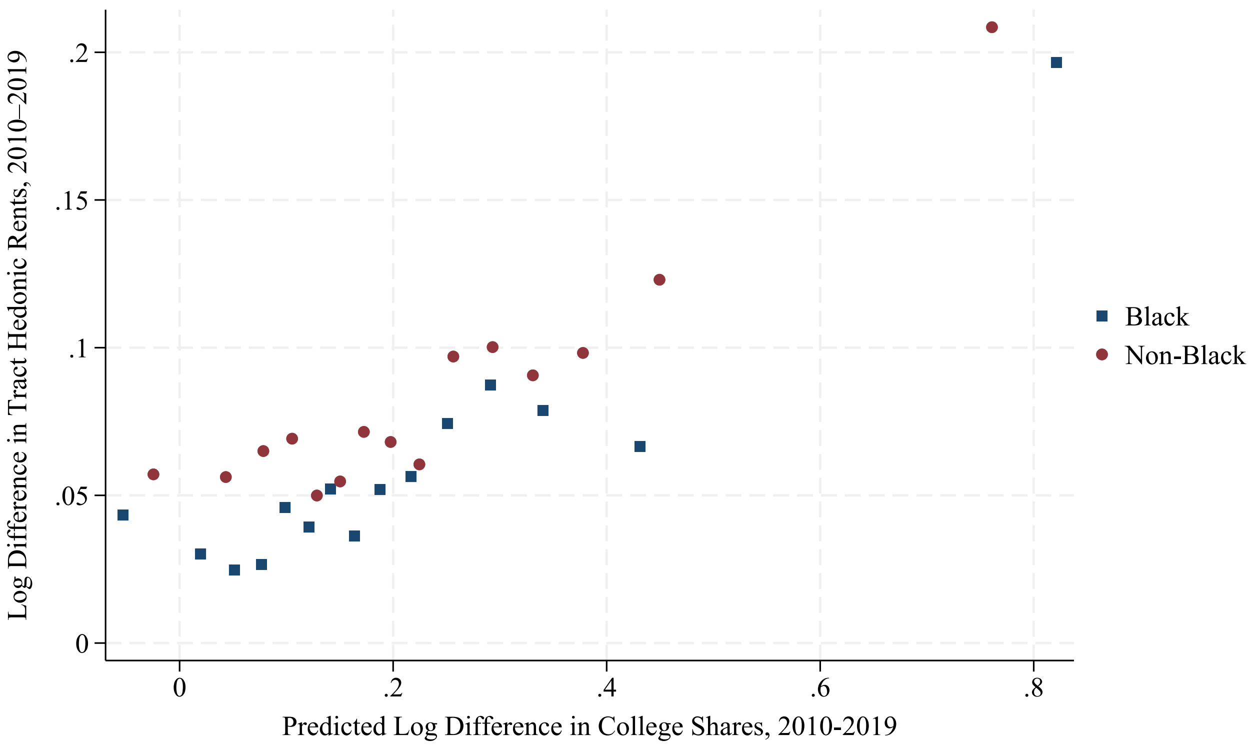
Notes: Table A15 reports coefficient estimates for our gentrification measure from regression equation **??**. The dependent variable is an indicator equal to one if the household resides in a different census tract in 2019 than in 2010. We instrument for gentrification using changes in the adult college-educated population, as described in Section **??**. Panel A replaces the gentrification measure with either (i) its counterpart constructed using only 5-year ACS aggregates, or (ii) the log-change in a tract's share of adults who are both white and hold at least a bachelor's degree. Panel B reports estimates from regression specifications with additional sample selection criteria discussed in the main text of Appendix **??**. Panel C reports estimates from regression specifications with additional regression controls also discussed in the main text of Appendix **??**. Otherwise, the coefficient estimates are analogous to those in Table A14.

Table A16. Gentrification and Experienced Neighborhood Characteristics

	(1) Baseline		(2) Gentrifying Tracts		(3) Rent-Controlled Tracts		(4) Long-Term Renters		(5) Movers		(6) Stayers	
	Non-Black	Black	Non-Black	Black	Non-Black	Black	Non-Black	Black	Non-Black	Black	Non-Black	Black
Panel A: College Share												
Coefficient estimate	0.508	0.427	0.542	0.462	0.340	0.311	0.595	0.526	0.160	0.093	1.000	1.000
Standard error	(0.017)	(0.015)	(0.021)	(0.018)	(0.015)	(0.016)	(0.014)	(0.016)	(0.016)	(0.014)	(0.000)	(0.000)
Kleibergen-Paap F-stat	[94.46]	[406.1]	[90.06]	[432.9]	[33.34]	[184.5]	[149.3]	[466.8]	[88.67]	[395.1]	[75.46]	[269.9]
Baseline controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
CBSA fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Panel B: White Share												
Coefficient estimate	0.069	0.098	0.078	0.097	0.041	0.044	0.082	0.185	0.002	-0.099	0.182	0.494
Standard error	(0.011)	(0.025)	(0.013)	(0.030)	(0.009)	(0.034)	(0.013)	(0.030)	(0.011)	(0.028)	(0.019)	(0.038)
Kleibergen-Paap F-stat	[94.46]	[406.1]	[90.06]	[432.9]	[33.34]	[184.5]	[149.3]	[466.8]	[88.67]	[395.1]	[75.46]	[269.9]
Baseline controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
CBSA fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Panel C: Hedonic Rents												
Coefficient estimate	0.135	0.121	0.158	0.146	0.058	0.058	0.161	0.144	0.033	0.031	0.258	0.239
Standard error	(0.012)	(0.011)	(0.014)	(0.014)	(0.008)	(0.009)	(0.014)	(0.014)	(0.005)	(0.006)	(0.024)	(0.022)
Kleibergen-Paap F-stat	[94.46]	[406.1]	[90.06]	[432.9]	[33.34]	[184.5]	[149.3]	[466.8]	[88.67]	[395.1]	[75.46]	[269.9]
Baseline controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
CBSA fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Panel D: Residential Market Access												
Coefficient estimate	0.034	0.023	0.036	0.026	0.012	0.011	0.039	0.021	0.030	0.024	0.050	0.021
Standard error	(0.006)	(0.006)	(0.007)	(0.007)	(0.005)	(0.005)	(0.007)	(0.007)	(0.005)	(0.004)	(0.009)	(0.010)
Kleibergen-Paap F-stat	[94.46]	[406.1]	[90.06]	[432.9]	[33.34]	[184.5]	[149.3]	[466.8]	[88.67]	[395.1]	[75.46]	[269.9]
Baseline controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
CBSA fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
N	392,000	278,000	324,000	219,000	212,000	190,000	177,000	114,000	249,000	190,000	143,000	87,500

Notes: Table A16 reports coefficient estimates for our gentrification measure from regression equation **??**. In Panel A, the dependent variable is the log difference in the share of adults with a college degree between each household's 2019 and 2010 census tracts. In Panel B, the dependent variable is the log difference in the share of white adults between each household's 2019 and 2010 census tracts. In Panel C, the dependent variable is the log difference in hedonic rents between each household's 2019 and 2010 census tracts. In Panel D, the dependent variable is the log difference in residential market access between each household's 2019 and 2010 census tracts. We instrument for gentrification using changes in the adult college-educated population, as described in Section **??**. The sample comprises urban low-income renter households observed in both 2010 and 2019 to live in the same CBSA, defined after applying restriction (h.ii) in Table A1. Otherwise, the coefficient estimates are analogous to those in Table A14.

FIGURE A6. Relationship Between Rents and Gentrification



Notes: Figure A6 shows the log change in tract-level hedonic rents as a function of the predicted log change in the share of college-educated adults in sample households' 2010 residential tracts. The points plot conditional means estimated from a regression that includes CBSA fixed effects (Cattaneo et al. 2024). Tracts are weighted separately by the number of Black and non-Black sample household residing in each tract in 2010. The samples comprise urban low-income renter households in 2010, defined after applying restriction (h.i) in Table A1.

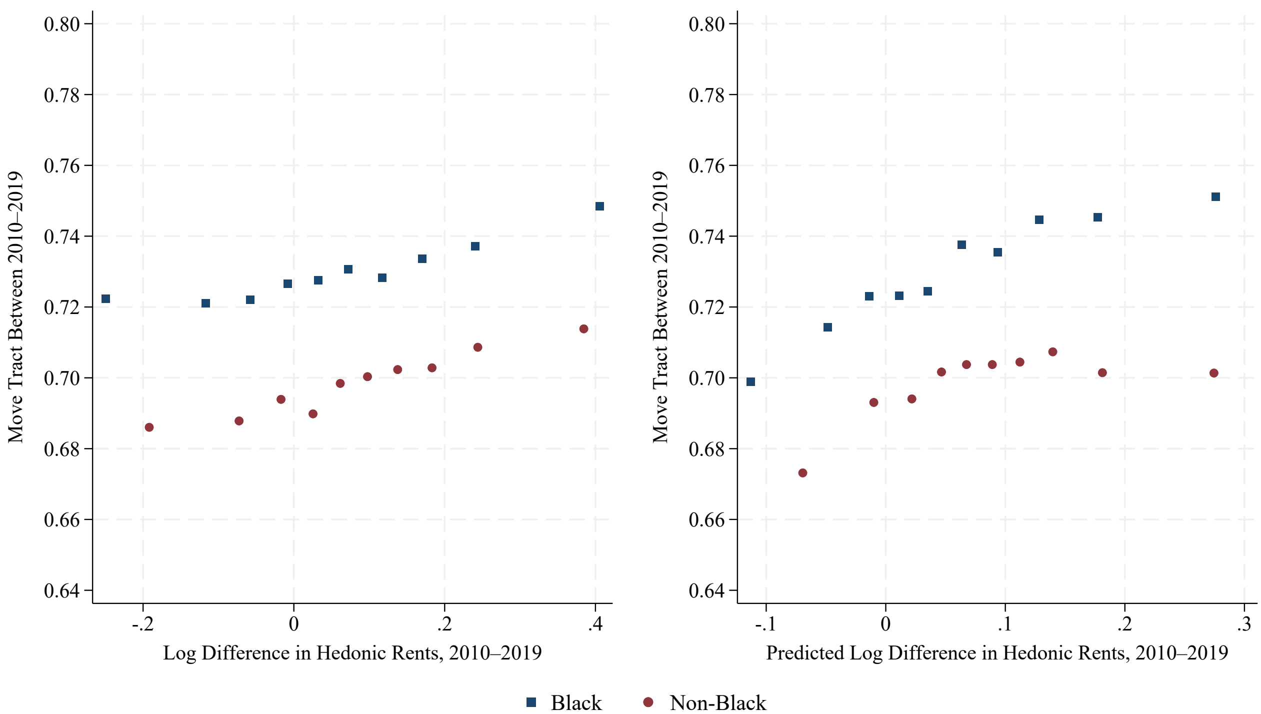
Relationship Between Rents and Gentrification.

Relationship Between Rents and Residential Mobility.

Gentrification and Market Access.

D.4. Gentrification and Economic Outcomes

Figure A7. Changing Rents and Residential Mobility



Notes: Figure A7 shows the probability incumbent renters move census tracts between 2010-2019 as a function of their origin tracts' log change in hedonic rents. The points plot conditional means estimated from regressions analogous to equation **??**, which include CBSA fixed effects and the full set of baseline controls listed in Table **??** (Cattaneo et al. 2024). The conditional means in the right-hand-side plot are constructed for log changes in hedonic rents predicted by changes in the adult college-educated population, as described in Section **??**. The sample comprises urban low-income renter households observed in both 2010 and 2019, defined after applying restriction (h.ii) in Table A1.

Table A17. Changing Rents and Residential Mobility

	(1) Baseline		(2) Gentrifying Tracts		(3) Rent-Controlled Tracts		(4) Long-Term Renters		(5) Movers	
	Non-Black	Black	Non-Black	Black	Non-Black	Black	Non-Black	Black	Non-Black	Black
Panel A: Moved Tract OLS										
Coefficient estimate	0.054	0.044	0.054	0.045	0.045	0.026	0.040	0.039	—	—
Standard error	(0.008)	(0.008)	(0.009)	(0.009)	(0.010)	(0.009)	(0.011)	(0.013)	—	—
Baseline controls	✓	✓	✓	✓	✓	✓	✓	✓		
CBSA fixed effects	✓	✓	✓	✓	✓	✓	✓	✓		
Panel B: Moved Tract IV										
Coefficient estimate	0.062	0.120	0.070	0.114	0.103	0.091	0.038	0.113	—	—
Standard error	(0.030)	(0.032)	(0.032)	(0.033)	(0.040)	(0.037)	(0.037)	(0.046)	—	—
Kleibergen-Paap F-stat	[201.5]	[188.3]	[173.0]	[165.6]	[102.9]	[99.25]	[194.4]	[142.5]	—	—
Baseline controls	✓	✓	✓	✓	✓	✓	✓	✓		
CBSA fixed effects	✓	✓	✓	✓	✓	✓	✓	✓		
Panel C: Moved CBSA IV										
Coefficient estimate	0.053	0.027	0.053	0.014	0.148	0.059	0.001	-0.004	—	—
Standard error	(0.019)	(0.018)	(0.019)	(0.018)	(0.046)	(0.024)	(0.021)	(0.020)	—	—
Kleibergen-Paap F-stat	[201.5]	[188.3]	[173.0]	[165.6]	[102.9]	[99.25]	[194.4]	[142.5]	—	—
Baseline controls	✓	✓	✓	✓	✓	✓	✓	✓		
CBSA fixed effects	✓	✓	✓	✓	✓	✓	✓	✓		
N	472,000	324,000	389,000	255,000	260,000	219,000	201,000	127,000	—	—
Panel D: Distance from Origin Tract IV										
Coefficient estimate	0.327	0.359	0.350	0.349	0.477	0.301	0.195	0.234	0.445	0.201
Standard error	(0.068)	(0.074)	(0.071)	(0.077)	(0.106)	(0.100)	(0.074)	(0.093)	(0.071)	(0.073)
Kleibergen-Paap F-stat	[190.0]	[183.7]	[163.2]	[160.6]	[92.35]	[96.05]	[194.9]	[139.8]	[197.6]	[182.3]
Baseline controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
CBSA fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Panel E: Distance from CBD IV										
Coefficient estimate	0.052	0.154	0.115	0.196	0.280	0.141	-0.074	0.053	0.052	0.179
Standard error	(0.080)	(0.064)	(0.079)	(0.064)	(0.162)	(0.104)	(0.061)	(0.055)	(0.123)	(0.097)
Kleibergen-Paap F-stat	[190.0]	[183.7]	[163.2]	[160.6]	[92.35]	[96.05]	[194.9]	[139.8]	[197.6]	[182.3]
Baseline controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
CBSA fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
N	392,000	278,000	324,000	219,000	212,000	190,000	177,000	114,000	249,000	190,000

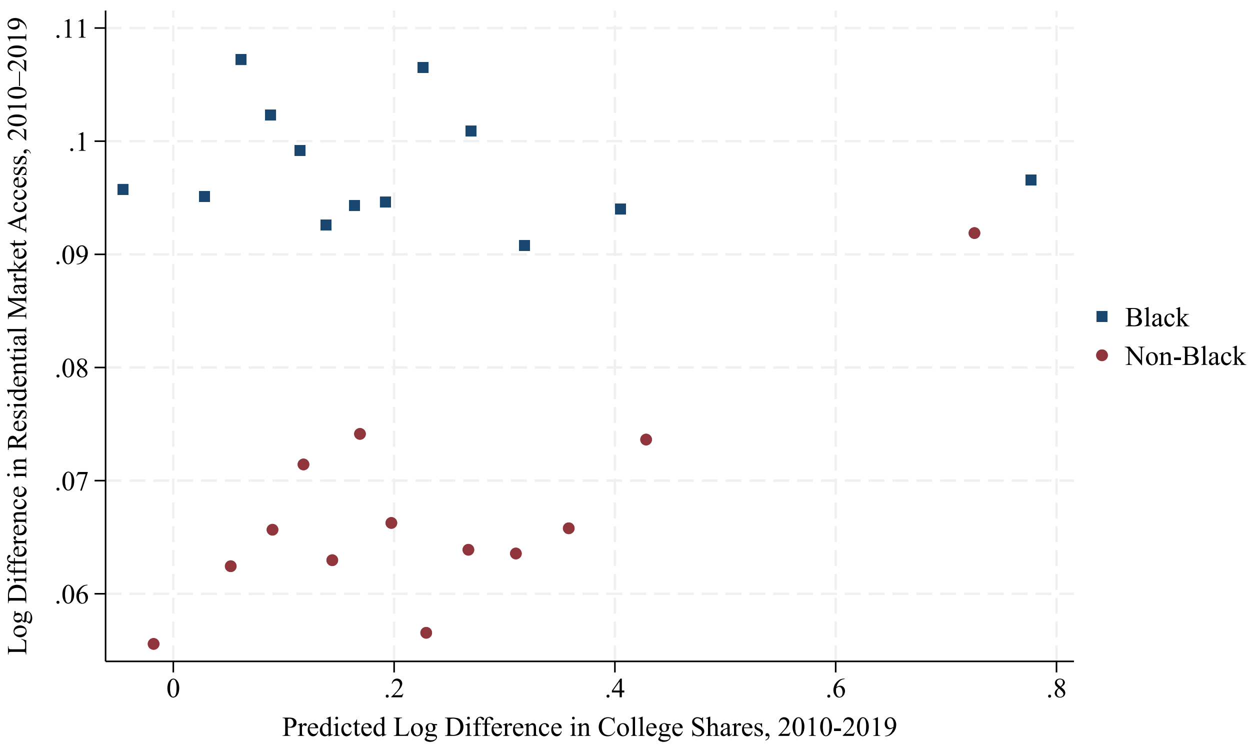
Notes: Table A17 reports coefficient estimates for log changes in tract-level hedonic rents from regression **??**, but with this hedonic rent measure replacing of our gentrification measure. In Panels A and B, the dependent variable is an indicator equal to one if the household resides in a different census tract in 2019 than in 2010. In Panel C, the dependent variable is an indicator equal to one if the household resides in a different CBSA in 2019 than in 2010. In Panel D, the dependent variable is one plus the log of the straight-line distance in miles from the centroid of the household's 2019 tract to the centroid of its 2010 origin tract. In Panel E, the dependent variable is one plus the log change in the household's distance in miles from its Metropolitan Division CBD to its tract centroid between 2010 and 2019. In Panels B-E, we instrument for log changes in hedonic rents using changes in the adult college-educated population, as described in Section **??**. All specifications include CBSA fixed effects and the full set of baseline controls listed in Table **??**. We define gentrifying tracts as those whose college share rose from 2010 to 2019; rent-controlled tracts as those covered by a rent-control ordinance in 2010; long-term renters as households residing in the same tract since at least 2005; and movers as households observed in a different census tract in 2019 than in 2010. Robust standard errors clustered at the origin census-tract level are in parentheses (Abadie et al. 2023). Kleibergen-Paap F-statistics are in square brackets. The sample comprises urban low-income renter households observed in both 2010 and 2019, defined after applying restriction (h.ii) in Table A1. Panels D and E further subset to households who are observed to reside in the same CBSA in 2019 as in 2010. All coeficient estimates and standard errors are rounded to three decimal places, while Kleibergen-Paap F-statistics are rounded to four significant figures.

Table A18. Changing Rents and Residential Mobility Heterogeneity

	Non-Black			Black			Baseline Controls	CBSA FEs
	Coefficient	F-Stat	N	Coefficient	F-Stat	N		
Panel A: Additional Sample Restrictions								
Inner City	0.042 (0.034)	134.6	293,000	0.076 (0.036)	119.6	209,000	✓	✓
No College Educated	0.058 (0.031)	196.9	425,000	0.133 (0.032)	189.5	293,000	✓	✓
Drop Outliers	0.095 (0.044)	108.4	434,000	0.269 (0.066)	62.35	282,000	✓	✓
Below Median College Share	0.042 (0.024)	310.5	795,000	0.065 (0.023)	268.6	483,000	✓	✓
Older Households	0.033 (0.045)	196.7	243,000	0.140 (0.040)	182.2	156,000	✓	✓
Panel B: Additional Controls								
Gentrification 2005-2009	0.061 (0.031)	183.2	472,000	0.121 (0.032)	180.3	324,000	✓	✓
Gentrification 2005-2009, Long-Term Renters	0.036 (0.038)	178.0	201,000	0.113 (0.047)	134.6	127,000	✓	✓
Log Change in Adult Population	0.055 (0.032)	162.6	472,000	0.134 (0.036)	150.9	324,000	✓	✓
Log Change in Tradable Market Access	0.065 (0.030)	200.5	472,000	0.118 (0.031)	189.1	324,000	✓	✓
Log Change in Racial Shares	0.068 (0.033)	187.0	472,000	0.101 (0.035)	161.1	324,000	✓	✓

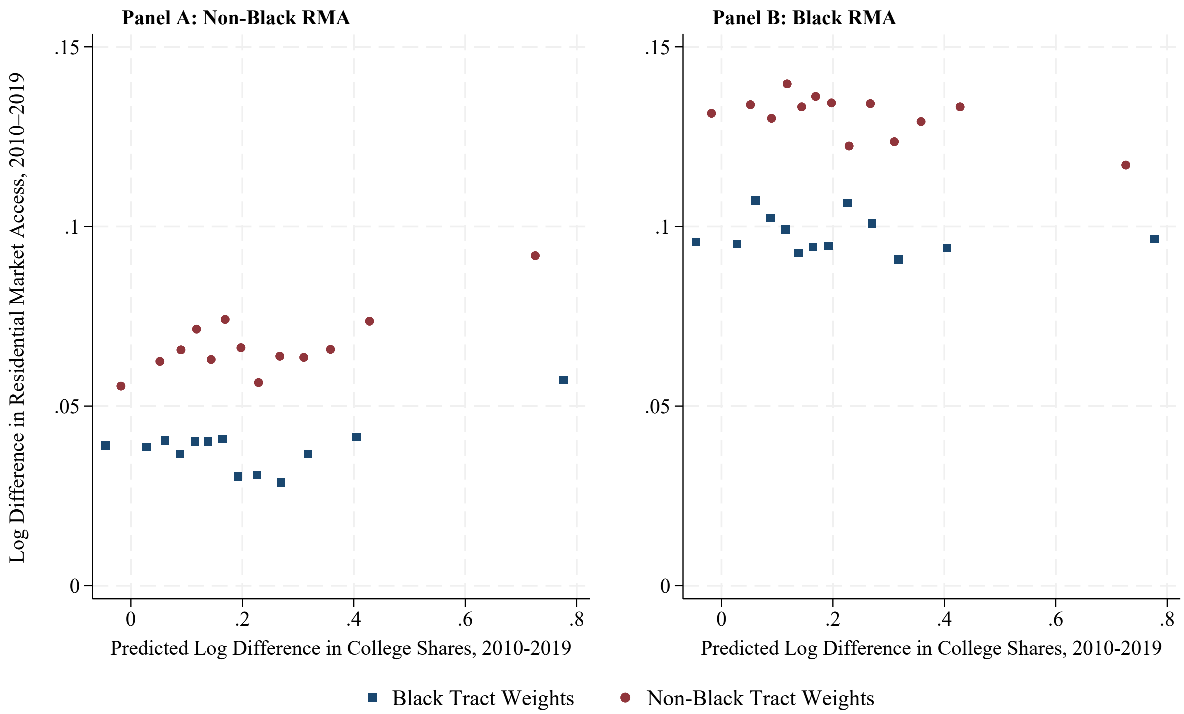
Notes: Table A18 reports coefficient estimates for log changes in tract-level hedonic rents from regression **??**, but with this hedonic rent measure replacing of our gentrification measure. The dependent variable is an indicator equal to one if the household resides in a different census tract in 2019 than in 2010. We instrument for log changes in tract-level hedonic rents using changes in the adult college-educated population, as described in Section **??**. Panel A reports estimates from regression specifications with additional sample selection criteria discussed in the main text of Appendix **??**. Panel B reports estimates from regression specifications with additional regression controls also discussed in the main text of Appendix **??**. All specifications include CBSA fixed effects and the full set of baseline controls listed in Table **??**. Robust standard errors clustered at the origin census-tract level are reported in parentheses (Abadie et al. 2023) alongside Kleibergen–Paap F-statistics. The baseline sample comprises urban low-income renter households observed in both 2010 and 2019, defined after applying restriction (h.ii) in Table A1. All coefficient estimates and standard errors are rounded to three decimal places, while Kleibergen-Paap F-statistics are rounded to four significant figures.

Figure A8. Gentrification and Residential Market Access



Notes: Figure A8 shows the log change in tract-level and race-specific residential market access as a function of the predicted log change in the share of college-educated adults in sample households' 2010 residential tracts. The points plot conditional means estimated from regressions that include CBSA fixed effects (Cattaneo et al. 2024). The residential market access measures are weighted by the number sample households from the corresponding race residing in each tract in 2010. The weights comprise urban low-income renter households in 2010, defined after applying restriction (h.i) in Table A1. The residential market access measures are derived from all low-income households in 2010 and 2019, defined after applying restriction (f) in Table A1.

Figure A9. Gentrification and Residential Market Access Weighted by Incumbent Shares



Notes: Figure A9 shows the log change in tract-level residential market access as a function of the predicted log change in the share of college-educated adults in sample households' 2010 residential tracts. Panel A plots log changes in residential market access derived from only non-Black low-income households in 2010 and 2019. Panel B plots log changes in residential market access derived from only Black low-income households in 2010 and 2019. The series within each panel are differentiated by the race of households comprising the tract-level weights. Otherwise, the figures are analogous to Figure A8.

Table A19. Gentrification and Economic Outcomes

	(1) Baseline		(2) Gentrifying Tracts		(3) Rent-Controlled Tracts		(4) Long-Term Renters		(5) Movers		(6) Stayers	
	Non-Black	Black	Non-Black	Black	Non-Black	Black	Non-Black	Black	Non-Black	Black	Non-Black	Black
Panel A: Non-Participation												
Coefficient estimate	-0.000	0.007	0.001	0.010	-0.000	0.007	-0.006	0.011	0.002	0.007	-0.002	0.009
Standard error	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.005)	(0.006)
Kleibergen-Paap F-stat	[94.46]	[406.1]	[90.06]	[432.9]	[33.34]	[184.5]	[149.3]	[466.8]	[88.67]	[395.1]	[75.46]	[269.9]
Baseline controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
CBSA fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Panel B: Income												
Coefficient estimate	0.028	-0.081	0.011	-0.109	0.006	-0.084	0.092	-0.151	-0.018	-0.100	0.073	-0.075
Standard error	(0.035)	(0.035)	(0.042)	(0.044)	(0.045)	(0.043)	(0.049)	(0.051)	(0.040)	(0.040)	(0.057)	(0.065)
Kleibergen-Paap F-stat	[94.46]	[406.1]	[90.06]	[432.9]	[33.34]	[184.5]	[149.3]	[466.8]	[88.67]	[395.1]	[75.46]	[269.9]
Baseline controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
CBSA fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
N	392,000	278,000	324,000	219,000	212,000	190,000	177,000	114,000	249,000	190,000	143,000	87,500
Panel C: Commute Time												
Coefficient estimate	0.016	0.003	0.028	0.011	-0.005	-0.015	0.021	0.016	0.018	-0.001	0.011	0.009
Standard error	(0.007)	(0.007)	(0.008)	(0.009)	(0.009)	(0.008)	(0.009)	(0.009)	(0.010)	(0.008)	(0.010)	(0.011)
Kleibergen-Paap F-stat	[84.27]	[377.1]	[81.87]	[390.2]	[36.24]	[179.0]	[125.4]	[455.7]	[113.5]	[412.6]	[57.11]	[193.1]
Baseline controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
CBSA fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
N	210,000	165,000	176,000	131,000	111,000	111,000	96,500	68,500	131,000	112,000	78,500	53,000

Notes: Table A19 reports coefficient estimates for our gentrification measure from regression equation **??**. In Panel A, the dependent variable is an indicator equal to one if we do not observe the household head with positive income in the LEHD in 2019. In Panel B, the dependent variable is the log difference in each household head's income (2010 dollars, plus one) between 2019 and 2010. In Panel C, the dependent variable is the log difference in each household head's commute time between 2019 and 2010. We instrument for gentrification using changes in the adult college-educated population, as described in Section **??**. In Panels A and B, the sample comprises urban low-income renter households we observe in both 2010 and 2019 to live in the same CBSA, defined after applying restriction (h.ii) in Table A1. In Panel C, the sample is further restricted to household heads we observe with positive income in the LEHD in both 2010 and 2019. Otherwise, the coefficient estimates are analogous to those in Table A14 and Table A16.

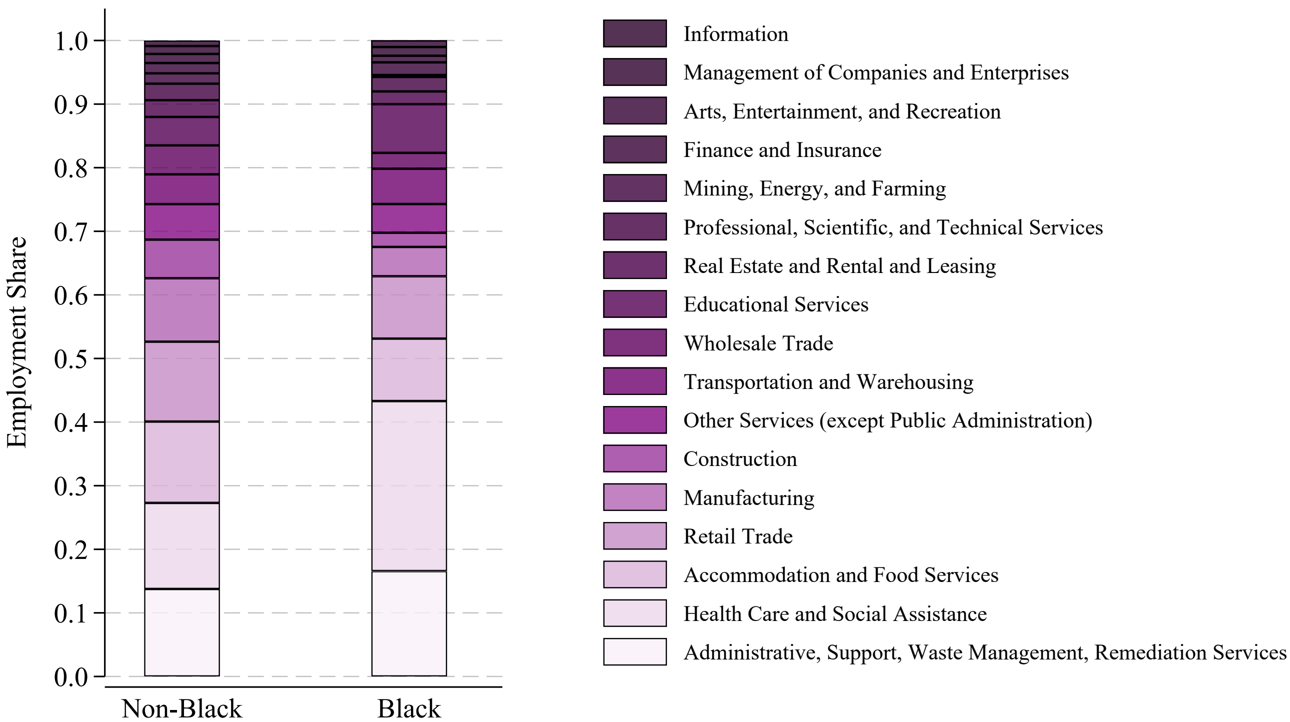
D.5. Baseline Job Characteristics and Mobility

Table A20. Baseline Job Characteristics and Mobility

	(1) Urban Low-Inc Renters		(2) Low-Inc Renters		(3) All Low-Inc HHs		(4) All HHs
	Non-Black	Black	Non-Black	Black	Non-Black	Black	
Panel A							
Switch job in 2010	0.2338	0.2492	0.2329	0.2477	0.2025	0.2325	0.1505
Number job switches 2010-2019	2.076	2.355	2.039	2.314	1.835	2.205	1.502
Standard deviation	(1.613)	(1.749)	(1.590)	(1.731)	(1.579)	(1.734)	(1.501)
Pseudo-median	[2.000]	[2.000]	[2.000]	[2.000]	[2.000]	[2.000]	[1.000]
Skew	⟨0.8165⟩	⟨0.6668⟩	⟨0.8419⟩	⟨0.6899⟩	⟨0.9657⟩	⟨0.7513⟩	⟨1.141⟩
Panel B							
Tradable job in 2010	0.4044	0.4512	0.3879	0.4461	0.3785	0.438	0.3892
Commute time in 2010	25.12	30.12	24.04	30.34	23.05	29.01	25.6
Standard deviation	(21.13)	(24.16)	(19.70)	(24.86)	(19.04)	(23.53)	(19.83)
Pseudo-median	[20.00]	[20.00]	[20.00]	[20.00]	[20.00]	[20.00]	[20.00]
Skew	⟨2.904⟩	⟨2.225⟩	⟨2.869⟩	⟨2.309⟩	⟨3.398⟩	⟨2.542⟩	⟨3.061⟩
Commute distance in 2010	9.783	9.611	10.05	10.09	10.96	10.86	13.36
Standard deviation	(9.551)	(9.185)	(9.829)	(9.380)	(10.21)	(9.812)	(11.45)
Pseudo-median	[7.252]	[7.309]	[7.384]	[7.798]	[8.287]	[8.311]	[10.51]
Skew	⟨2.626⟩	⟨3.715⟩	⟨2.435⟩	⟨2.807⟩	⟨2.617⟩	⟨2.368⟩	⟨2.084⟩
N	517,000	351,000	1,994,000	799,000	5,866,000	1,594,000	24,830,000

Notes: Table A20 summarizes household heads' job mobility and baseline job characteristics in 2010. To align with our annual residential-tract assignments, each household head is assigned one firm ID per calendar year, determined by the establishment from which they earned the highest wage income. Panel A reports job switching either between 2010 and 2011 or year-to-year over 2010-2019. By construction, a household head can switch at most once per year, and job-mobility statistics are computed only when a household head is assigned a job in both the beginning- and end-year associated with the statistic. Panel B summarizes characteristics of household heads' assigned jobs in 2010. A job is classified as tradable if its industry's trade costs fall in the bottom 75% of industries, as defined in Gervais and Jensen (2019) (see Footnote 28 for details). Commute-time and commute-distance statistics are computed for household heads matched to the 2010 person-level ACS. Commute time is measured in minutes rounded to the nearest five, and distance is measured in miles. Columns (1)–(4) use the samples defined after applying restrictions (h.i), (g), (f), and (d) in Table A1. Household counts correspond to Panel A.

FIGURE A10. Sectoral Employment Shares



Notes: Figure A10 plots sectoral employment shares for urban low-income renter households in 2010, defined after applying restriction (h.i) in Table A1. Industries are stacked in ascending order of the non-Black sample's employment shares. Sectors aggregate 2-digit NAICS (2012) industries as follows: 23–Construction; 11, 21, and 22–Mining, Energy, and Farming; 31, 32, 33–Manufacturing; 42–Wholesale Trade; 44, 45–Retail Trade; 48, 49–Transportation and Warehousing; 52–Finance and Insurance; 51–Information; 53–Real Estate and Rental and Leasing; 54–Professional, Scientific, and Technical Services; 55–Management of Companies and Enterprises; 56–Administrative and Support and Waste Management and Remediation Services; 61–Educational Services; 62–Health Care and Social Assistance; 71–Arts, Entertainment, and Recreation; 72–Accommodation and Food Services; and 81–Other Services (except Public Administration).

D.6. Welfare Regressions

Appendix E. Welfare Simulations

E.1. Empirical Setup

Data Construction. We conduct our counterfactual welfare simulations using data procurred outside the Census restricted IRE. This lets us report tract-level estimates of expected welfare that would otherwise fail the Census Bureau's disclosure requirements for small-area geographic estimates. We construct tract-level aggregates in a way that closely matches those we construct inside the Census restricted IRE. We compute raw tract-level measures of median gross rent paid as well as the share of adult residents over 25 years old with and without a college degree from the 5-year ACS tract estimates (Manson et al. 2022). The US Census Bureau does not publicly release tract-level counts of low-income Black and non-Black renters separately by low ($\bar{\tau} = 1$) and high ($\bar{\tau} = 2$) tenure. We therefore approximate these counts in several steps. First, we construct tract-level counts of Black and non-Black renters in each tract from the 5-year ACS tract estimates. Second, we approximate tract-level counts of low-income Black and non-Black households from 5-year ACS income-bin tabulations. We define low-income households as those earning among the bottom tercile of their respective CBSA's income distribution for cohorts of 35 to 45 year olds, estimated from the Integrated Public Use Microdata Series (IPUMs) with person-level weights (Steven Ruggles and Williams 2199). Third, we construct Public Use Microdata Area (PUMA) estimates of counts of low-income renter households by race and length of tenure from IPUMS using household weights. We finally combine these estimates by scaling the tract-level counts of low-income Black and non-Black households by the share of Black and non-Black households who are renters in each tract, and then apportioning the resulting counts into low- and high-tenure groups according to the share of low-income renters of each race in the corresponding PUMA who are low- and high-tenure.[62]

We similarly construct tract-level and race-specific measures of residential market access for low-income households using publicly available data to approximate our estimates inside the Census IRE. We estimate tract-level jobs belonging to low-income household heads by race using the LEHD Origin-Destination Employment Statistics (LODES) workplace-area job counts by tract, year, and industry sector, excluding public administration jobs. We allocate these sector jobs to Black and non-Black low-income household heads using baseline race-specific sector shares from our analysis sample recorded in Table A10, rescaling each job count by the ratio of the sector's baseline share of low-income renter employment to its share of total LODES employment in each year. We then sum across sectors to approximate tract-year race-specific low-income employment counts up to a common scale factor. We combine these job counts with the tract-level counts of low-income Black and non-Black households and solve the corresponding unique (to scale) fixed point in **??** to obtain values for RMA_{nt}^{k} in each tract and year for which the LODES job counts are available. When constructing these measures, we use the 50$^{\text{th}}$ percentile 2010 parameter cluster estimates (η_{2010}^{kc}) reported in Table A5. We also calibrate tract-pair commute times based off tract pairs' physical distance and their distances to the metropolitan division CBD, ensuring they are consistent with households' mean commute times reported in Table 1 and estimates from our commute-time forecasts in Table A4.

We crosswalk all our tract-level estimates to 2010 tabulation Census tract boundaries using household population weights, and deflate gross rents and low-income cutoffs to real 2010 dollars. We impute the rents for high-tenure households using... We approximate aggregate tenure transition probabilities... Finally, we smooth each tract-level estimate using a kernel-weighted average of estimates from nearby tracts in the same metropolitan division and from nearby years (inclusive of the base year), where the weights decay with both geographic distance and time.[63]

Tract Aggregate Forecasting. We compute our expected welfare measures by backward induction from a terminal period in which we assume the economy is in steady state. Ideally, this terminal period would be far enough in the future that discounting makes the steady-state assumption largely irrelevant for

[62] We experimented with apportioning the PUMA-level counts of low-income renters by tenure level to tracts based on the tract estimates, but we found the overall count of low-income renter households in each PUMA to be particularly variable across years. We therefore only rely on the ratio of high- and low-tenure low-income renter households in each PUMA-year.

[63] Our counterfactual welfare analyses require smooth tract-level estimates. Otherwise, households in our simulations could arbitrage measurement noise by repeatedly moving to neighborhoods with artificially low rents or spuriously high market access and amenities.

our expected welfare measures. Because we estimate tract-level aggregates only through 2021, setting a later terminal year requires forecasting tract characteristics forward. We use 2030 as the terminal year because our forecasts of tract characteristics appear credible through that horizon, and because the present-discounted contribution of the terminal steady-state assumption becomes quite small. After some experimentation, we used the following tract-fixed-effects panel vector error correction model (VECM) to forecast neighborhood characteristics through 2030.

We define the log and the log-first-difference of each endogenous neighborood characteristic Y_{nt}, as $y_{nt} \equiv \log(Y_{nt})$ and $\Delta y_{nt} \equiv \ln\left(\frac{Y_{nt}}{Y_{nt-1}}\right)$, respectively.[64] We estimate the cointegrating equation using all endogenous tract characteristics in 2011-2021 and tract fixed effects (ζ_n), defining the error-correction term (ECT) as the within-tract residual,

$$\mathrm{ECT}_{nt} = y_{nt}^{\mathrm{rent}} - \hat{\theta}' \boldsymbol{y}_{nt}^{-\mathrm{rent}} - \hat{\zeta}_n,$$

where y_{nt}^{rent} is the log of median gross rents in tract n and year t, and $\mathbf{y}_{nt}^{-\mathrm{rent}}$ stacks the logs of the remaining endogenous tract characteristics. We use ECT_{nt} to estimate the following VECM,

$$\Delta \boldsymbol{y}_{nt} = \mathbf{A}\Delta \boldsymbol{y}_{nt-1} + \boldsymbol{\phi}\mathrm{ECT}_{nt-1} + \upsilon_n + \boldsymbol{u}_{nt},$$

where $\mathbf{A}$ is the one-period VAR coefficient matrix and $\boldsymbol{\phi}$ collects the error-correction coefficients. We forecast through 2022-2030 recursively by generating one-step-ahead predictions for the log of each endogenous characteristic and corresponding deviations from the cointegrating relationship,[65]

$$\hat{\boldsymbol{y}}_{nt} = \hat{\boldsymbol{y}}_{nt-1} + \widehat{\Delta \boldsymbol{y}_{nt}}, \qquad \widehat{\mathrm{ECT}}_{nt} = \hat{y}_{nt}^{\mathrm{rent}} - \hat{\theta}' \hat{\boldsymbol{y}}_{nt}^{-\mathrm{rent}} - \hat{\zeta}_n.$$

When transforming back to levels, we replace any negative endogenous values with the smallest positive estimate within the same CBSA-year.

E.2. Counterfactual Simulation Details

For each census tract in our analysis, we compute and compare two tract-level measures of expected welfare for low-income Black and non-Black renter households. Both tract-level measures are constructed from the perspective of the year 2010, separately for each tenure length $\bar{\tau}$. The first expected welfare measure assumes the city's economy is in a steady-state in the year 2010. We denote this measure $\bar{V}_{SS}^{k}\left(x(n,\bar{\tau}), \bar{\omega}_{2010}^{kc}\right)$. We compare it to an alternative expected welfare measure, $\bar{V}_{Gent}^{k}\left(x(n,\bar{\tau}), \bar{\omega}_{2010}^{kc}\right)$, that assumes households believe their city's tract characteristics will evolve just as they do in the data.[66] We quantify the welfare effects of gentrification for incumbent low-income renter households by comparing

[64] The endogenous neighborhood characteristics we forecast forward are counts of adults over 25 years with and without a college degree, median gross rents, counts of both Black and non-Black low-income renter households with low tenure ($\bar{\tau} = 1$), and separate residential market access measures for Black and non-Black low-income households.

[65] Since many tracts have very low values for Black low-income renter counts, we winsorize that variable's one-step-ahead log-difference forecasts at the 35th and 60th percentiles of its in-sample log-difference distribution. For all other variables, we winsorize one-step-ahead log-difference forecasts at the corresponding 5th and 95th percentiles.

[66] This assumption implies that households' expected ex-ante continuation values equal their realized counterparts in the simulations. Maybe discuss this a little more.

these tract-level measures across tracts that gentrify and those that do not.

Steady-State Expected Welfare. We compute $\bar{V}^k_{SS}\left(x(n,\bar{\tau}), \bar{\omega}^{kc}_{2010}\right)$ by holding observed city-level state variables fixed at their 2010 values and then solving for time-invariant continuation values and exogenous amenities $\left\{\tilde{\xi}^k_{n,2010}\right\}_n$ that generate a stationary distribution of type-k low-income renter households. To define a stationary distribution, we denote $q^k_{nt}(\bar{\tau})$ as the model-implied share of each city's type-k households with tenure length $\bar{\tau}$ living in neighborhood n in period t. Stacking $q^k_{nt}(\bar{\tau})$ yields a $2(N^c+1)$-dimensional vector $\boldsymbol{q}^k_t$. We define a stationary distribution in year t among our sample of low-income renter households as $\boldsymbol{q}^k_t = \boldsymbol{\Lambda}^k_t \boldsymbol{q}^k_t$, where $\boldsymbol{\Lambda}^k_t$ is a transition probability matrix constructed so the distribution of type-k household tract shares evolve such that,

$$q^k_{nt}(\bar{\tau}) = \sum_{\bar{\tau}'}\sum_{n'} q^k_{n't-1}(\bar{\tau}') \cdot p^k_n\left(x(n',\bar{\tau}'), \bar{\omega}^{kc}_t\right) \cdot f^x(x(n,\bar{\tau})|n, x(n',\bar{\tau}')) \quad \forall\, n, \bar{\tau}.$$

We define a steady-state equilibrium in 2010 as a stationary distribution among type-k low-income renter households in which continuation values are time-invariant and all city-specific state variables are held fixed at their 2010 levels for all $t > 2010$. The resulting time-invariant continuation values are our steady-state measures of expected welfare, $\bar{V}^k_{SS}\left(x(n,\bar{\tau}), \bar{\omega}^{kc}_{2010}\right)$. Algorithm 1 describes how we compute these welfare measures in practice, along with the corresponding vector of exogenous amenities.[67]

Expected Welfare Change. We compute $\bar{V}^k_{Gent}\left(x(n,\bar{\tau}), \bar{\omega}^{kc}_{2010}\right)$ by first computing steady-state expected welfare measures in the year 2030, $\bar{V}^k_{SS}\left(x(n,\bar{\tau}), \bar{\omega}^{kc}_{2030}\right)$. We compute these and 2030 exogenous amenities analogously to how we compute their 2010 counterparts, but hold observed city-level state variables fixed at their 2030 forecast values. We then solve for $\bar{V}^k_{Gent}\left(x(n,\bar{\tau}), \bar{\omega}^{kc}_{2010}\right)$ via backward induction given $\bar{V}^k_{SS}\left(x(n,\bar{\tau}), \bar{\omega}^{kc}_{2030}\right)$ until we obtain $2N^c$ tract-level measures of expected welfare from the perspective of 2010, one for each aggregate tenure level in each tract. During each step of the backward induction procedure, we solve for the levels of exogenous amenities that rationalize the observed share of type-k low-income renter households with low aggregate tenure across census tracts in that year. To do so, we repeat steps 5 and 6 in Algorithm 1, but substitute $\left\{\bar{V}^k_{SS}\left(x(n,\bar{\tau}), \bar{\omega}^{kc}_t\right)\right\}_{n,\bar{\tau}}$ for the measures of expected welfare computed in the previous year's iteration.

With $\bar{V}^k_{Gent}\left(x(n,\bar{\tau}), \bar{\omega}^{kc}_{2010}\right)$ and $\bar{V}^k_{SS}\left(x(n,\bar{\tau}), \bar{\omega}^{kc}_{2010}\right)$ in hand, we define expected welfare change as,

$$\Delta\bar{V}^k\left(x(n,\bar{\tau}), \bar{\omega}^{kc}_{2010}\right) \equiv \bar{V}^k_{Gent}\left(x(n,\bar{\tau}), \bar{\omega}^{kc}_{2010}\right) - \bar{V}^k_{SS}\left(x(n,\bar{\tau}), \bar{\omega}^{kc}_{2010}\right). \tag{A14}$$

Welfare Effect Decomposition. We decompose $\Delta\bar{V}^k\left(x(n,\bar{\tau}), \bar{\omega}^{kc}_{2010}\right)$ into one component capturing incumbent renters' own-tract flow utilities and another capturing their option value. To do so, we apply the Hotz and Miller (1993) inversion to $\bar{V}^k_{Gent}\left(x(n,\bar{\tau}), \bar{\omega}^{kc}_{2010}\right)$ and $\bar{V}^k_{SS}\left(x(n,\bar{\tau}), \bar{\omega}^{kc}_{2010}\right)$, taking the household's

[67] In step 6 of Algorithm 1, we exploit that exogenous amenities do not vary with households' length of tenure. Although Algorithm 1 need not generate an exactly stationary distribution for type-k households with high neighborhood tenure ($\bar{\tau} = 2$), we find that it closely approximates stationarity in practice. We define $\hat{q}^k_{nt}(\bar{\tau}=1)$ as the empirical counterpart to $q^k_{nt}(\bar{\tau}=1)$.

Algorithm 1 Compute Steady State Expected Welfare

1: Guess value of unobserved neighborhood amenities: $\left\{\tilde{\xi}_n^k\right\}_n$

Compute continuation values:

2: Guess $\left\{\bar{V}_{SS}^k\left(x(n,\bar{\tau}),\bar{\omega}_t^{kc}\right)\right\}_{n,\bar{\tau}}$

3: Compute for all n and $\bar{\tau}$,

$$\bar{V}_{SS}^{\prime k}\left(x(n,\bar{\tau}),\bar{\omega}_t^{kc}\right)=\ln\left(\sum_{n'\in\mathcal{N}^c}\exp\left(\bar{u}_{n'}^k\left(x(n,\bar{\tau}),\bar{\omega}_t^{kc}\right)+\delta\sum_{\bar{\tau}'}\bar{V}_{SS}^k\left(x(n',\bar{\tau}'),\bar{\omega}_t^{kc}\right)\cdot f^x(x(n',\bar{\tau}')|n',x(n,\bar{\tau}))\right)\right)+\gamma$$

4: Set $\bar{V}_{SS}^k\left(x(n,\bar{\tau}),\bar{\omega}_t^{kc}\right)=\bar{V}_{SS}^{\prime k}\left(x(n,\bar{\tau}),\bar{\omega}_t^{kc}\right),\quad\forall\, n,\bar{\tau}$

Repeat steps 3 - 4 until $\max_{n,\bar{\tau}}\left\{\left|\bar{V}_{SS}^k\left(x(n,\bar{\tau}),\bar{\omega}_t^{kc}\right)-\bar{V}_{SS}^{\prime k}\left(x(n,\bar{\tau}),\bar{\omega}_t^{kc}\right)\right|\right\}<\epsilon_V$ for some $\epsilon_V>0$.

Compute exogenous amenities:

5: Compute the probability a type-k household chooses n given $\left\{\bar{V}_{SS}^k\left(x(n,\bar{\tau}),\bar{\omega}_t^{kc}\right)\right\}_{n,\bar{\tau}}$ and $\left\{\tilde{\xi}_n^k\right\}_n$

$$p_n^k\left(x(\underline{n},\bar{\tau}),\bar{\omega}_t^{kc}\right)=\frac{\exp\left(v_n^k\left(x(\underline{n},\bar{\tau}),\bar{\omega}_t^{kc}\right)\right)}{\sum_{n'\in\mathcal{N}^c}\exp\left(v_{n'}^k\left(x(\underline{n},\bar{\tau}),\bar{\omega}_t^{kc}\right)\right)},\quad\forall\, n,\underline{n},\bar{\tau}$$

6: Update exogenous amenities using observed neighborhood shares of low-income Black and non-Black households with low tenure, $q_n^k(\bar{\tau}=1)$:

$$\tilde{\xi}_n^{\prime k}=\tilde{\xi}_n^k+\ln\left(\hat{q}_n^k(\bar{\tau}=1)\right)-\ln\left(\sum_{\bar{\tau}'}\sum_{\underline{n}}q_{\underline{n}}^k(\bar{\tau}')\cdot p_n^k\left(x(\underline{n},\bar{\tau}'),\bar{\omega}_t^{kc}\right)\cdot f^x(x(n,\bar{\tau}=1)|n,x(\underline{n},\bar{\tau}'))\right) \tag{A13}$$

Set $\tilde{\xi}_n^k=\tilde{\xi}_n^{\prime k}$.

Repeat steps 2 - 6 until $\max_n\left\{\left|\tilde{\xi}_n^k-\tilde{\xi}_n^{\prime k}\right|\right\}<\epsilon_\xi$ for some $\epsilon_\xi>0$.

Take $\bar{V}_{SS}^k\left(x(n,\bar{\tau}),\bar{\omega}_t^{kc}\right)$ as the measures of steady-state expected welfare with the city-specific state variables fixed at their period t levels.

origin tract as the reference option in the inversion:

$$\begin{aligned}\bar{V}_{\mathcal{U}}^k\left(x(n,\bar{\tau}),\bar{\omega}_{2010}^{kc}\right)&=\ln\left(\sum_{n'\in\mathcal{N}^c}\exp\left(v_{\mathcal{U}n'}^k\left(x(n,\bar{\tau}),\bar{\omega}_{2010}^{kc}\right)\right)\right)+\gamma\\&=v_{\mathcal{U}n}^k\left(x(n,\bar{\tau}),\bar{\omega}_{2010}^{kc}\right)+\ln\left(1+\sum_{n'\neq n}\exp\left(v_{\mathcal{U}n'}^k\left(x(n,\bar{\tau}),\bar{\omega}_{2010}^{kc}\right)-v_{\mathcal{U}n}^k\left(x(n,\bar{\tau}),\bar{\omega}_{2010}^{kc}\right)\right)\right)+\gamma\\&=\underbrace{v_{\mathcal{U}n}^k\left(x(n,\bar{\tau}),\bar{\omega}_{2010}^{kc}\right)}_{\text{payoff from staying}}-\underbrace{\ln\left(p_{\mathcal{U}n}^k\left(x(n,\bar{\tau}),\bar{\omega}_{2010}^{kc}\right)\right)}_{\text{option value of moving}}+\gamma,\end{aligned}$$

where $\mathcal{U}\in\{SS,Gent\}$ indicates that the respective object can be computed within the steady-state environment or under the observed evolution of tract characteristics. The second equality highlights how the probability of staying in one's origin neighborhood reflects the relative desirability of a household's broader choice set. When other tracts are less desirable relative to the origin tract, households are less likely to leave, and expected welfare lies closer to the value of remaining in the origin tract. In the limiting case in which the household stays in its origin tract with probability one, the option value of moving is zero, so expected welfare is equal to the origin tract's conditional value function, $\bar{V}^k_{\mathfrak{U}}\left(x(n,\bar{\tau}), \bar{\omega}^{kc}_{2010}\right) = v^k_{\mathfrak{U}n}\left(x(n,\bar{\tau}), \bar{\omega}^{kc}_{2010}\right)$.

We can express expected welfare change under this decomposition as,

$$
\begin{aligned}
\Delta\bar{V}^k\left(x(n,\bar{\tau}), \bar{\omega}^{kc}_{2010}\right) &\equiv \bar{V}^k_{Gent}\left(x(n,\bar{\tau}), \bar{\omega}^{kc}_{2010}\right) - \bar{V}^k_{SS}\left(x(n,\bar{\tau}), \bar{\omega}^{kc}_{2010}\right) \\
&= \underbrace{\Delta\, v^k_n\left(x(n,\bar{\tau}), \bar{\omega}^{kc}_{2010}\right)}_{\Delta \text{ payoff from staying}} - \underbrace{\Delta \ln\left(p^k_n\left(x(n,\bar{\tau}), \bar{\omega}^{kc}_{2010}\right)\right)}_{\Delta \text{ option value of moving}}, \qquad \forall\, n,\ \bar{\tau}. \qquad \text{(A15)}
\end{aligned}
$$

where $\Delta v^k_n\left(x(n,\bar{\tau}), \bar{\omega}^{kc}_{2010}\right)$ and $\Delta \ln\left(p^k_n\left(x(n,\bar{\tau}), \bar{\omega}^{kc}_{2010}\right)\right)$ are the differences in origin tracts' conditional value functions and option values computed under the observed evolution of tract characteristics and under the steady-state environment. Recursively substituting the conditional value function (7) into A15 yields,

$$
\begin{aligned}
\Delta\bar{V}^k\left(x(n,\bar{\tau}), \bar{\omega}^{kc}_{2010}\right) = &\sum_{s=0}^{20}\delta^s \sum_{\bar{\tau}_{2010+s}} \Pi_{2010+s}(\bar{\tau}_{2010+s}|\bar{\tau}) \left[\Delta\bar{u}^k_n\left(x(n,\bar{\tau}_{2010+s}), \bar{\omega}^{kc}_{2010+s}\right) + \Delta \ln\left(p^k_n\left(x(n,\bar{\tau}_{2010+s}), \bar{\omega}^{kc}_{2010+s}\right)\right)\right] \\
&+ \frac{\delta^{21}}{1-\delta}\sum_{\bar{\tau}_{2031}} \Pi_{2031}(\bar{\tau}_{2031}|\bar{\tau}) \left[\Delta\bar{u}^k_n\left(x(n,\bar{\tau}_{2031}), \bar{\omega}^{kc}_{2031}\right) + \Delta \ln\left(p^k_n\left(x(n,\bar{\tau}_{2031}), \bar{\omega}^{kc}_{2031}\right)\right)\right],
\end{aligned}
$$

where $\Pi_t(\bar{\tau}_t|\bar{\tau})$ is the probability of being in tenure state $\bar{\tau}_t$ after staying in tract n until period t given the initial state $\bar{\tau}$,

$$
\Pi_{t+1}(\bar{\tau}_{t+1}|\bar{\tau}) = \sum_{\bar{\tau}_t} f^x(x(n,\bar{\tau}_{t+1})|n, x(n,\bar{\tau}_t)) \cdot \Pi_t(\bar{\tau}_t|\bar{\tau}).
$$

Because flow utility is linear in tracts' characteristics, we can additionally decompose differences in the payoffs from staying in n during any period t,

$$
\Delta\bar{u}^k_n\left(x(n,\bar{\tau}_t), \bar{\omega}^{kc}_t\right) = \tilde{\beta}^k_I \Delta \ln\left(\bar{I}^k_{nt}\right) + \tilde{\beta}^k_r \Delta \log\left(r_{\bar{\tau}_t nt}\right) + \tilde{\beta}^k_A \Delta \ln\left(\frac{\text{Coll}_{nt}}{\text{Pop}_{nt}}\right) + \Delta\xi^k_{nt},
$$

where Δ now denotes the difference between the period-t components of the payoff from staying in tract n under the observed evolution of tract characteristics and under the steady-state environment. Finally, we can denote the present discounted value for any component of the expected welfare change decomposition as,

$$
\text{PDV}\Delta\chi = \sum_{s=0}^{20}\delta^s \sum_{\bar{\tau}_{2010+s}} \Pi_{2010+s}(\bar{\tau}_{2010+s}|\bar{\tau})\Delta\chi_s + \frac{\delta^{21}}{1-\delta}\sum_{\bar{\tau}_{2031}} \Pi_{2031}(\bar{\tau}_{2031}|\bar{\tau})\Delta\chi_{2031},
$$

for all $\chi_s \in \left\{\ln\left(p^k_n\left(x(n,\bar{\tau}_{2010+s}), \bar{\omega}^{kc}_{2010+s}\right)\right),\ \bar{u}^k_n\left(x(n,\bar{\tau}_{2010+s}), \bar{\omega}^{kc}_{2010+s}\right),\ \tilde{\beta}^k_r \ln\left(r_{\bar{\tau}_s ns}\right),\ \tilde{\beta}^k_A \ln\left(\frac{\text{Coll}_{ns}}{\text{Pop}_{ns}}\right),\ \tilde{\beta}^k_I \ln\left(\bar{I}^k_{ns}\right),\ \xi^k_{ns}\right\}$ and analogously for χ_{2031}. This setup lets us write the final decompositions we present in Section **??**:

$$\Delta \bar{V}^k \left(x(n,\bar{\tau}), \bar{\omega}^{kc}_{2010}\right) = \underbrace{\mathrm{PDV}\Delta \bar{u}^k_n \left(x(n,\bar{\tau}), \bar{\omega}^{kc}\right)}_{\text{origin-tract flow utility changes}} - \underbrace{\mathrm{PDV}\ln\left(p^k_n\left(x(n,\bar{\tau}), \bar{\omega}^{kc}\right)\right)}_{\text{option value changes}} \tag{A16}$$

and,

$$\begin{aligned} \Delta \bar{V}^k \left(x(n,\bar{\tau}), \bar{\omega}^{kc}_{2010}\right) = & \underbrace{\mathrm{PDV}\tilde{\beta}^k_I \Delta \ln\left(\bar{I}^k_n\right)}_{\text{market access changes}} + \underbrace{\mathrm{PDV}\tilde{\beta}^k_r \Delta \ln\left(r_{\bar{\tau}n}\right)}_{\text{rent changes}} + \underbrace{\mathrm{PDV}\tilde{\beta}^k_A \Delta \ln\left(\frac{\mathrm{Coll}_n}{\mathrm{Pop}_n}\right)}_{\text{amenity changes}} \\ & + \underbrace{\mathrm{PDV}\Delta \xi^k_n}_{\text{exogenous changes}} - \underbrace{\mathrm{PDV}\ln\left(p^k_n\left(x(n,\bar{\tau}), \bar{\omega}^{kc}\right)\right)}_{\text{option value changes}}. \end{aligned} \tag{A17}$$

Calibrating Model Parameters. Discussion of minor model calibrations coming soon.

E.3. Additional Results

Table A21. Gentrification and Experienced Welfare

	(1) Baseline		(2) Rent-Controlled Tracts		(3) Long-Term Renters	
	Non-Black	Black	Non-Black	Black	Non-Black	Black
Panel A: Baseline						
Coefficient estimate	0.081	-0.116	0.055	-0.205	0.125	-0.125
Standard error	(0.054)	(0.069)	(0.067)	(0.075)	(0.069)	(0.091)
Kleibergen-Paap F-stat	[98.21]	[445.2]	[37.41]	[224.0]	[143.5]	[515.4]
Baseline controls	✓	✓	✓	✓	✓	✓
CBSA fixed effects	✓	✓	✓	✓	✓	✓
Panel b: Exclude Hedonic Rents						
Coefficient estimate	0.417	0.110	0.311	0.024	0.502	0.105
Standard error	(0.057)	(0.059)	(0.074)	(0.056)	(0.075)	(0.078)
Kleibergen-Paap F-stat	[98.21]	[445.2]	[37.41]	[224.0]	[143.5]	[515.4]
Baseline controls	✓	✓	✓	✓	✓	✓
CBSA fixed effects	✓	✓	✓	✓	✓	✓
Panel C: Exclude Market Access						
Coefficient estimate	0.010	-0.236	0.012	-0.255	0.054	-0.222
Standard error	(0.059)	(0.064)	(0.071)	(0.070)	(0.073)	(0.084)
Kleibergen-Paap F-stat	[98.21]	[445.2]	[37.41]	[224.0]	[143.5]	[515.4]
Baseline controls	✓	✓	✓	✓	✓	✓
CBSA fixed effects	✓	✓	✓	✓	✓	✓
Panel D: Exclude College Shares						
Coefficient estimate	-0.834	-0.383	-0.602	-0.431	-0.894	-0.433
Standard error	(0.067)	(0.068)	(0.069)	(0.083)	(0.077)	(0.089)
Kleibergen-Paap F-stat	[98.21]	[445.2]	[37.41]	[224.0]	[143.5]	[515.4]
Baseline controls	✓	✓	✓	✓	✓	✓
CBSA fixed effects	✓	✓	✓	✓	✓	✓
Panel E: Exclude Tenure Length						
Coefficient estimate	0.082	-0.109	0.060	-0.199	0.122	-0.120
Standard error	(0.055)	(0.069)	(0.068)	(0.075)	(0.070)	(0.091)
Kleibergen-Paap F-stat	[98.21]	[445.2]	[37.41]	[224.0]	[143.5]	[515.4]
Baseline controls	✓	✓	✓	✓	✓	✓
CBSA fixed effects	✓	✓	✓	✓	✓	✓
Panel F: Exclude Moving Costs						
Coefficient estimate	0.174	-0.025	0.202	-0.123	0.128	-0.050
Standard error	(0.077)	(0.077)	(0.095)	(0.086)	(0.088)	(0.097)
Kleibergen-Paap F-stat	[98.21]	[445.2]	[37.41]	[224.0]	[143.5]	[515.4]
Baseline controls	✓	✓	✓	✓	✓	✓
CBSA fixed effects	✓	✓	✓	✓	✓	✓
N	408,000	269,000	214,000	171,000	178,000	109,000

Notes: Table A21 reports the estimated coefficients on our gentrification measure from regression **??**. In Panel A, the dependent variable is experienced welfare. The dependent variables in the remaining Panels are constructed identically but omit a given component of experienced utility by setting the corresponding welfare coefficient to zero when computing experienced utility. We instrument for gentrification using changes in the adult college-educated population, as described in Section **??**. All specifications include CBSA-by-final-sample-year fixed effects and the full set of baseline controls listed in Table **??**. The specifications additionally include all remaining neighborhood characteristics that comprise households' 2010 flow utilities net of moving costs and value for residential tenure. Namely, the natural logarithms of their 2010 tract's college share and hedonic rent, as well as the tract's estimated exogenous amenities in 2010. Robust standard errors clustered at the origin census-tract level are in parentheses (Abadie et al. 2023). The sample comprises urban, low-income renter households initially residing in gentrifying tracts, defined after applying restriction (h.iii) in Table A1.