Preferences over the Racial Composition of Neighborhoods: Estimates and Implications*

Morris A. Davis
Rutgers University
mdavis@business.rutgers.edu

Jesse Gregory
University of Wisconsin and NBER
jmgregory@ssc.wisc.edu

Daniel A. Hartley
Federal Reserve Bank of Chicago
Daniel.A.Hartley@chi.frb.org

December 2, 2024

Abstract

Using data from 197 metro areas, we estimate the parameters of a dynamic, forward-looking neighborhood choice model where households have preferences over the racial composition of the neighborhood in which they live. Using multiple metro areas in the estimation sample enables us to develop a new, shift-share IV strategy to estimate the impact of the racial composition of neighborhoods on location choice that relies only on across-metro comparisons of similarly situated neighborhoods. Our neighborhood-level instrument is constructed by interacting national-average, across-neighborhood sorting patterns with respect to neighborhood-level topography with metro-level shares of households by demographic subgroup. For a given configuration of neighborhood-level topographic data, the instrument predicts variation in neighborhood-level racial shares that is attributable exclusively to variation in metro-level shares of demographic subgroups. We find that households in many different demographic subgroups have strong preferences to live in neighborhoods consisting mostly or entirely of households of the same race. These preferences are sufficiently strong that model simulations suggest that the current demographic composition of neighborhoods is not stable.

*This paper was previously circulated with the title "The Long-Run Effects of Low-Income Housing on Neighborhood Composition." We thank seminar participants and discussants over the years and appreciate detailed comments from Milena Almagro, Rebecca Diamond, Andra Ghent and Ed Glaeser. We also gratefully acknowledge financial support from the National Science Foundation (grant SES-1629422). The views expressed herein are those of the authors and do not necessarily represent those of the Federal Reserve Bank of Chicago or the Federal Reserve System.

1 Introduction

In this paper, we propose a new method for estimating household preferences for racial diversity in the neighborhoods in which they live. As documented by Cutler, Glaeser, and Vigdor (1999) and others, households in the United States continue to be racially segregated. This segregation may be the outcome of some or many households having explicit preferences over the racial mix of the neighborhood in which they live. An alternative explanation is that households do not directly care about the racial mix of their neighborhood, but segregation occurs because households of different races differentially value or differ in their ability to pay for various neighborhood amenities. Of course, both of these forces may be present and contributing to the segregation of neighborhoods today, especially once historical policies and patterns are taken into consideration.

There are important policy implications of understanding why segregation occurs. If households do not have strong preferences regarding the race of their neighbors, then government programs that incentivize racial integration may be able to increase neighborhood racial diversity. If some households have preferences exhibiting "homophily," the desire to live in neighborhoods with households of their same race, then government programs may shift people around in the short run, but may not increase racial diversity in the long run as households respond to the program and to other households moving.

To get at these important questions, we specify a dynamic, discrete-choice model of within-metro neighborhood choice and estimate household preferences for the racial mix of their neighborhood using a new shift-share strategy (Bartik, 1991). We use a dynamic model in order to discuss the short- and long-run impacts of hypothetical government policies on the demographic composition of neighborhoods. Our shift-share research design requires estimating a different version of the dynamic location choice model specifically tailored for each of a large number of metro areas, because the approach for identifying racial preferences involves comparisons of similarly situated neighborhoods across metro areas. For that reason, the geographic scope of our analysis needs to be substantially larger than existing studies that estimate dynamic discrete choice models of neighborhood choice using data from a single metro area. We apply our research design using panel data on neighborhood choices in nearly 200 metro areas from the New York Fed/Equifax Consumer Credit Panel and detailed topographic data in each neighborhood in these metros, a key input to our shift-share IV.

We estimate the model's parameters using a two-step, BLP-style procedure (Berry, Levinsohn, and Pakes, 1995; Bayer, Ferreira, and McMillan, 2007), appropriately adapted for our dynamic modeling environment (Bishop, 2012; Davis, Gregory, Hartley, and Tan, 2021). In the first step, we use maximum likelihood to obtain estimates of the indirect flow utility that each census tract ("neighborhood") provides to households from each of 54 different demographic subgroups ("types") in each metro area. In the second step, we regress these first-stage estimates of neighborhood indirect utilities on variables describing neighborhood racial composition, instrumenting for the racial share variables which are likely to be correlated with neighborhood-level unobserved amenities.

We construct the IVs for neighborhood racial shares using a shift-share strategy that harnesses the fact that topography inside a metro area is highly predictive of where households with different demographic and socio-economic characteristics tend to live (Ye and Becker, 2018; Heblich, Trew, and Zylberberg, 2021; Ye and Becker, 2024) and that metro areas vary substantially in their overall demographic structure. The shift-share instrument that we generate for racial composition in every neighborhood in every metro interacts (a) national average neighborhood-choice patterns with respect to topographic variables for each of the 54 demographic types in our data, computed using prediction equations that we pre-estimate using a national sample of neighborhoods, with (b) metro-level shares of those household types. Thus, for a given configuration of topographic data, the instrument predicts variation in neighborhood-level racial shares that is attributable exclusively to variation in metro-level shares of demographic types. In estimation, we control directly for the impact of topography on preferences, such that this across-metro variation in our instrument (exclusively determined by variation in metro-level type shares) is the only source of identification of the impact of racial shares on preferences.

The recent literature on the econometrics of shift-share instrumental variables has emphasized that either exogeneity of the shares (Goldsmith-Pinkham, Sorkin, and Swift, 2020) or exogeneity of the shifts (Borusyak, Hull, and Jaravel, 2022), but not neccessarily both, is required for shift-share IVs to satisfy instrument exogeneity. Our approach relies on an assumption of "share exogeneity," specifically the metro-level type shares, and importantly we do not require topography to be exogenous to neighborhood amenities. To verify this theoretical result, we perform Monte Carlo simulation experiments that demonstrate our approach produces consistent estimates of preference parameters in an environment where households directly sort on amenities (including topography) when making location choices.

Our estimates suggest that some, but not all, types of households have relatively strong preferences for neighborhood racial composition exhibiting homophily. The fact that households do not uniformly wish to live in more segregated neighborhoods hints that there may be some room for policy to engineer more racially integrated neighborhoods than currently observed. We use our model to predict the outcomes of one such hypothetical policy that, absent any resorting response, should mechanically increase racial integration relative to current data. We simulate a policy shock in which all current low-income housing tax credit developments expand by 10 percent and the new units are populated by low-income households with the same demographic mix as the metro average for low-income households. We then compute a new steady state after the policy is implemented and all households have had a chance to resort in response to this policy and the location-choice decisions of other households.

In every metro that we simulate, the new steady state features much more racial segregation (relative to the starting point of the simulations) due to the endogenous resorting of the population that takes place over time. This resorting and the increase in segregation that results occurs because many types of households have strong preferences exhibiting homophily. When we repeat simulations of the hypothetical policy shock, but after significantly "shrinking" (but not eliminating) estimates of racial preferences for all types of households, the steady state that emerges in many metro areas is more racially integrated than the starting point of the simulations.

We also use simulations of the model to understand the relationship between household expectations and the speed at which households resort in response to implementation of this policy. If households are myopic and assume current racial shares will persist indefinitely (only updating expectations when observed shares change), then it can take decades for households to resort in response to the policy.[1] In contrast, if households expect a new steady state to emerge immediately after the policy is implemented, they act rapidly and the new steady state in fact typically emerges within 10 years.

Our paper contributes to established literatures on how the racial composition of neighborhoods affects location choices, and how this impacts the equilibrium allocation of people across neighborhoods. Kuminoff, Smith, and Timmins (2013) provides a survey of equilibrium models with endogenous location choices, including a few papers where households have preferences over the racial composition of neighborhoods, for example Bayer, Ferreira, and McMillan (2007). Focusing on estimation, some prominent recent papers have used quasi-experimental variation in geographically proximate locations to estimate how the racial composition of neighborhoods affects location choice, for example Almagro, Chyn, and Stuart (2022) and Bayer, Casey, McCartney, Orellana-Li, and Zhang (2022).[2] Almagro, Chyn, and Stuart (2022) use randomness in the timing of demolitions of public housing in Chicago, under the assumption that other amenities are held fixed over time, and Bayer,

[1]In the classic paper by Schelling (1971), households are assumed to solve a sequence of static models when making decisions, implying expectations are myopic over future neighborhood composition.

[2]These papers are in the spirit of Bayer, Ferreira, and McMillan (2007), who use boundaries for school attendance zones in San Francisco to generate plausibly exogenous changes in neighborhood racial composition, under the assumption that unobserved amenities on both sides of the boundary are roughly the same.

Casey, McCartney, Orellana-Li, and Zhang (2022) focus on changes to the racial composition of neighbors that are geographically close (on their block) and therefore plausibly share the same bundle of amenities not related to race. Despite our use of a different model, different data, and an estimation procedure that relies on a different source of variation for identification, our results about preferences are in line with findings of Almagro, Chyn, and Stuart (2022) and Bayer, Casey, McCartney, Orellana-Li, and Zhang (2022) and others: Many households have preferences that exhibit homophily.[3]

Our work also directly relates to other design-based studies estimating White households' migration response to inflows of Black households. Boustan (2010), who studies post-WWII White migration from central cities to suburbs, uses a shift-share approach similar to that of Card (2001) to estimate the impact of Black inflows on White migration.[4] Shertzer and Walsh (2019) use an IV approach similar to Boustan (2010), but with a prediction equation that generates within-city White migration in response to Black migration between 1900 and 1930.[5]

Finally, there is an extensive literature on the dynamics of neighborhood change. See Ellen (2000) for an overview and Ellen and Torrats-Espinosa (2019) for a discussion of racial change in the context of gentrification. An important recent study in this literature is Caetano and Maheshri (2021), which exploits the time-series structure of a dynamic location-choice decision model to develop instruments based on sufficiently lagged data that identifies exogenous components of the current racial composition of neighborhoods in San Francisco. Similar to our findings, Caetano and Maheshri (2021) estimate that household preferences exhibit homophily. Given these preferences, their model predicts that the racial composition of neighborhoods in San Francisco will change over time.

[3] Additionally, in line with our findings, Aliprantis, Carroll, and Young (2022) find that preferences (homophily) rather than wealth explain differences in the socio-economic status of the neighborhoods in which Black and White households reside. Christensen and Timmins (2021) and Christensen and Timmins (2022) also show that steering and barriers to entry may also play a role in determining the neighborhoods to which Black households have access.

[4] In related papers, Derenoncourt (2022) studies the impact of southern Black migration to northern and western commuting zones on inter-generational income mobility of households in those zones and Shi, Hartley, Mazumder, and Rajan (2022) study the impact of this migration on urban renewal projects at the city level. For identification, Boustan (2010), Derenoncourt (2022) and Shi, Hartley, Mazumder, and Rajan (2022) use southern state-level push factors interacted with historical county-level migration patterns.

[5] In Shertzer and Walsh (2019), the source of variation arises from differences in Black out-migration rates from southern states interacted with historical northern city neighborhood destinations of migrants from those states. Relative to Shertzer and Walsh (2019), we are using different shifts and shares and are estimating racial preferences within the context of a dynamic model of location choice.

2 A Simple Location-Choice Model (for Exposition)

To illustrate our new method of identifying preferences for neighborhood racial composition, we start by analyzing a simplified, static version of the dynamic location-choice model we eventually estimate. This simplified model makes clear the conditions required for identification of parameters. Additionally, we perform Monte Carlo simulation experiments using the simplified model to evaluate the properties of the estimator we propose under different assumptions about the data-generating process.

Relative to the full dynamic model that we ultimately estimate, there are two differences. First, we assume that when a household changes its location, the move is costless. This implies that households do not need to be forward looking when making location choices, and a household's location-choice decision can be described by a static model. Second, with respect to the demographic composition of any location, in this simplified model we assume household utility is linear over one demographic variable, the percentage of the location that is comprised of Black households. When we estimate the full dynamic model, we assume household utility in any location is a quadratic function over two demographic variables, the share of households in that location that are Black and the share that are Hispanic.

2.1 The Model

Demand for Locations. In any given metro area m, each household (indexed by i) must choose a location in that metro (indexed by ℓ) in which to live. We assume households can freely move to any location in the metro area. We also assume households are not allowed to exit from their current metro area.

Households belong to "types" that capture differences in demographics and socio-economic status. Each type of household has its own set of parameters that determines the expected utility of living in any location. Household i of type τ living in location ℓ in metro area m receives utility of

$$u^{\tau}_{i,\ell,m} \;=\; \delta^{\tau}_{\ell,m} \;+\; \epsilon^{\tau}_{i,\ell,m} \tag{1}$$

$\delta^{\tau}_{\ell,m}$ is the portion of utility that is common to all type τ households choosing location ℓ in metro m and $\epsilon^{\tau}_{i,\ell,m}$ is a shock that is specific to household i that is drawn iid from the Type I Extreme Value Distribution. The inclusion of $\epsilon^{\tau}_{i,\ell,m}$ ensures that not every household of the same type optimally chooses the same location.

$\delta^{\tau}_{\ell,m}$ is assumed to have three components: (1) the price of a rental unit in neighborhood ℓ in metro m, $r_{\ell,m}$; (2) the fraction of neighborhood ℓ in metro m that is comprised of Black households, $S^b_{\ell,m}$; (3) and "amenities" in neighborhood ℓ in metro m, $\mathcal{A}^\tau_{\ell,m}$. We specify $\delta^\tau_{\ell,m}$ as

$$\delta^\tau_{\ell,m} \;=\; -a^\tau_r \log r_{\ell,m} \;+\; a^\tau_1 S^b_{\ell,m} \;+\; \mathcal{A}^\tau_{\ell,m} \tag{2}$$

The parameters a^τ_r and a^τ_1 are allowed to vary across household types. The amenity value of neighborhood ℓ in metro m, $\mathcal{A}^\tau_{\ell,m}$, is also allowed to vary across types.

Households are assumed to know $r_{\ell,m}$ and $S^b_{\ell,m}$ in every ℓ in metro m.[6] Each household optimally chooses its utility-maximizing location ℓ in metro m after observing $\epsilon^\tau_{i,\ell,m}$ in all locations in the metro. Suppose there are J_m possible locations in metro m, and denote $\ell^*_{i,m}$ as the optimal location choice for household i in metro m. This choice satisfies

$$\ell^*_{i,m} \;=\; \underset{\ell=1,\ldots,J_m}{\operatorname{argmax}} \left\{u^\tau_{i,\ell,m}\right\}$$

Given $\epsilon^\tau_{i,\ell,m}$ is drawn iid from the Type 1 extreme value distribution, the probability that a household of type τ optimally chooses location k in metro m takes a multinomial logit form

$$p^\tau_{k,m} \;=\; \frac{\exp(\delta^\tau_{k,m})}{\sum_{k'=1}^{J_m} \exp(\delta^\tau_{k',m})} \tag{3}$$

and the probability that a household of type τ optimally chooses location k in metro m relative to the probability that that same household optimally chooses another location ℓ in metro m has the simple expression

$$\log\left(p^\tau_{k,m}/p^\tau_{\ell,m}\right) \;=\; \delta^\tau_{k,m} - \delta^\tau_{\ell,m} \tag{4}$$

where $p^\tau_{j,m}$ is the probability that a household of type τ chooses location j in metro m for $j = \ell, k$.

Supply of Housing. We assume each neighborhood has its own housing supply curve. In simulations of this version of the model that follow, we consider two cases for all locations: Perfectly elastic provision of housing at a fixed price and perfectly inelastic and fixed supply of housing in every location. In the full dynamic model that we estimate, we allow each neighborhood to have its own housing-supply elasticity that is taken from Baum-Snow and Han (2022).

[6] Formally, each household takes $r_{\ell,m}$ and $S^b_{\ell,m}$ as given when making decisions. In equilibrium, $r_{\ell,m}$ and $S^b_{\ell,m}$ must be consistent with the decisions all households have made.

2.2 Estimation with IV

We now discuss our strategy to estimate each type's preferences for the racial composition of neighborhoods. We construct IVs for neighborhood racial shares using a shift-share strategy that harnesses the fact that topography inside a metro area is highly predictive of where households with different demographic and socio-economic characteristics tend to live (Ye and Becker, 2018; Heblich, Trew, and Zylberberg, 2021; Ye and Becker, 2024). Topography may be correlated with household type because the natural amenities of various topographies may vary and the willingness to pay for those amenities is likely to vary in the population. Ye and Becker (2024) mention, specifically, that the estimated income elasticity of better scenery (for example) is greater than one and that "elevation variance" reduces access to public transit, which imposes costs on poor households that are less likely to own a car.

To create the shift-share instrument, we pool all the metro-area data and use a nationally-estimated relationship of within-metro "relative topography" (the value of any topographic variable after metro-area fixed effects have been removed) and neighborhood choice probabilities to predict the probability each type will occupy any given neighborhood. In each metro area, we combine these predicted probabilities with the observed distribution of household types to predict the share of households in each location that are Black. Thus, for a given configuration of topographic data (the direct effect of which we control for), the instrument predicts variation in neighborhood-level racial shares that is attributable exclusively to variation in metro-level shares of household types.

To be clear, we are not assuming topographic variables are instruments unrelated to amenities. Rather, we use topography to predict where households of various types will tend to live. This allows the instrument to exploit the fact that the impact of metro-level shares of certain types of households on predicted neighborhood-level racial shares varies by topography. We discuss this in great detail in sections 2.2.2 and 2.2.3.

2.2.1 Constructing the Estimator

To estimate the parameters of this model, we assume the following data are available:

1. *Type shares.* For each metro area, the total share of the population accounted for by each household type, denoted by s_m^τ for type τ in metro m.
2. *Neighborhood choice probabilities.* Estimates of the market share for each location ℓ in each metro area m, for many metro areas, and for all household types τ, $\hat{p}_{\ell,m}^\tau > 0$.
3. *Topographic data.* Topographic data for each location ℓ in each metro area m, $TOP_{\ell,m}$. In this section, $TOP_{\ell,m}$ has one dimension.
4. *Rental prices.* The constant-quality rental price of a housing unit in each location ℓ in

each metro m, $r_{\ell,m}$.

Note that when items (1) and (2) are appropriately combined, we can construct the share of Black households in each neighborhood ℓ in each metro area m as follows:

$$S^b_{\ell,m} \quad = \quad \frac{\sum\limits_{\tau'} \mathcal{I}\left(\tau' \in Black\right) s^{\tau'}_m \; \hat{p}^{\tau'}_{\ell,m}}{\sum\limits_{\tau} s^{\tau}_m \; \hat{p}^{\tau}_{\ell,m}} \tag{5}$$

In this equation τ and τ' are indexes for household type and $\mathcal{I}\left(\tau' \in Black\right)$ is an indicator function that is equal to 1 if type τ' households are Black, 0 otherwise. When the population of metro m is equal to 1.0, the numerator of (5) is the total population of Black households living in tract ℓ in metro m and the denominator is the total population of households living in tract ℓ in metro m.

BLP Stage 1: Given these data, our procedure to estimate type-specific parameters of this model has two steps as in Berry, Levinsohn, and Pakes (1995). First, we obtain estimates of $\delta^{\tau}_{\ell,m}$ for all ℓ, m, τ. Call these estimates $\hat{\delta}^{\tau}_{\ell,m}$. To do this, for each τ in each metro m we normalize $\hat{\delta}^{\tau}_{k,m} = 0$ for one particular location $k = 1$. Given this normalization, we use equation (4) to map data on $\hat{p}^{\tau}_{\ell,m}$ to $\hat{\delta}^{\tau}_{\ell,m}$ for all the $\ell \neq 1$ in metro m,

$$\forall \ell \neq 1, \qquad \hat{\delta}^{\tau}_{\ell,m} \quad = \quad \log\left(\hat{p}^{\tau}_{\ell,m} / \, \hat{p}^{\tau}_{1,m}\right)$$

The focus of our study is the impact of racial composition on neighborhood demand. In this simplistic model where only the Black share of households enters utility, that is the coefficient a^{τ}_1. In line with this focus, and consistent with our strategy to estimate the more complicated dynamic model, we assume we know from a previous study the parameter a^{τ}_r, the sensitivity of location choices to exogenously shifting rental prices. Bayer, McMillan, Murphy, and Timmins (2015) discuss why estimation of a^{τ}_r is an unusually difficult undertaking and make the case for bringing in outside evidence, as we do here.[7]

Given presumed knowledge of a^{τ}_r, we define a new variable:

$$\hat{d}^{\tau}_{\ell,m} \quad = \quad \hat{\delta}^{\tau}_{\ell,m} + a^{\tau}_r \log r_{\ell,m}$$

[7] When we estimate the more complicated dynamic model, we take a^{τ}_r from Davis, Gregory, Hartley, and Tan (2021). Out of concern that our results may be sensitive to our assumed values of a^{τ}_r, we confirmed that the parameter estimates on racial composition in the full dynamic model that we estimate are essentially unaffected by reasonable variation (doubling or halving) of a^{τ}_r. This result occurs because the estimated coefficients of racial composition on indirect utility are orders of magnitude more important than the coefficient on rental prices.

Then we can rewrite equation (2) as

$$\hat{d}^{\tau}_{\ell,m} \;\; = \;\; a_1^{\tau} S^b_{\ell,m} \; + \; A^{\tau}_{\ell,m} \tag{6}$$

Note the switch of notation for amenities, from $\mathcal{A}^{\tau}_{\ell,m}$ in equation (2) to $A^{\tau}_{\ell,m}$ in equation (6) to explicitly account for the fact that $\delta^{\tau}_{\ell,m}$ in equation (2) is replaced with $\hat{d}^{\tau}_{\ell,m}$ in equation (6).

BLP Stage 2, Constructing the Instrument: We generate a prediction equation for location choice in each metro for each household type that only depends on the topography of a location in that metro. For each type of household we pool all data across locations and metros and estimate

$$\log \hat{p}^{\tau}_{\ell,m} \;\; = \;\; \alpha^{\tau}_m \; + \; b^{\tau} TOP_{\ell,m} \; + \; \nu^{\tau}_{\ell,m} \tag{7}$$

where α^{τ}_m is a metro-area fixed effect that can vary by τ, b^{τ} is a coefficient that maps relative topography to location choice probabilities that can also vary by τ, and $\nu^{\tau}_{\ell,m}$ is a location, metro, and type-specific error.

Once (7) is estimated for all household types, we use it to predict the probability each type lives in any location in any metro area. Denote the predicted probability for type τ in location ℓ in metro m as $\hat{\hat{p}}^{\tau}_{\ell,m}$. We then create a predicted Black share in each location ℓ in each metro m, $Z^b_{\ell,m}$, as

$$Z^b_{\ell,m} \;\; = \;\; \frac{\sum_{\tau'} \mathcal{I}\left(\tau' \in Black\right) s^{\tau'}_m \; \hat{\hat{p}}^{\tau'}_{\ell,m}}{\sum_{\tau} s^{\tau}_m \; \hat{\hat{p}}^{\tau}_{\ell,m}} \tag{8}$$

Similar to equation (5), τ and τ' are indexes for household type and $\mathcal{I}\left(\tau' \in Black\right)$ is an indicator function that is equal to 1 if type τ' households are Black, 0 otherwise. The numerator of (8) is the *predicted* total population of Black households living in tract ℓ in metro m and the denominator is the *predicted* total population of households living in tract ℓ in metro m, where all of these predictions are based only on topographic data. Restated, we use equation (7) to predict where everyone will live based only on topographic data and then, given this prediction, use metro-level household-type shares to calculate the predicted share of each location that is comprised of Black households.

BLP Stage 2, IV: Once we have created $Z^b_{\ell,m}$, we use 2SLS to estimate a^{τ}_1. Specifically, we pool data across all locations in all metros and for each type of household we estimate

$$\hat{d}^{\tau}_{\ell,m} \; = \; \theta^{\tau}_{m} \; + \; a^{\tau}_{1} S^{b}_{\ell,m} \; + \; g^{\tau}\left(TOP_{\ell,m}\right) \; + \; \upsilon^{\tau}_{\ell,m} \tag{9}$$

where $\hat{d}^{\tau}_{\ell,m} = \hat{\delta}^{\tau}_{\ell,m} + a^{\tau}_{r} \log r_{\ell,m}$, θ^{τ}_{m} is a metro-area fixed effect that varies by τ, $g^{\tau}\left(TOP_{\ell,m}\right)$ is a flexible function of $TOP_{\ell,m}$, and $\upsilon^{\tau}_{\ell,m}$ is an error term. The first stage is

$$S^{b}_{\ell,m} \; = \; \vartheta_{m} \; + \; \gamma Z^{b}_{\ell,m} \; + \; g_{1}\left(TOP_{\ell,m}\right) \tag{10}$$

where ϑ_{m} is a metro-area fixed effect and $g_{1}\left(TOP_{\ell,m}\right)$ is a different flexible function of $TOP_{\ell,m}$.

2.2.2 Discussion of Variation in the Instrument

To see why the Black share of a neighborhood ℓ in metro m is an endogenous variable in equation (9), causing OLS to yield biased estimates of a^{τ}_{1}, we show equation (5) again here for convenience

$$S^{b}_{\ell,m} \; = \; \frac{\sum_{\tau'} \mathcal{I}\left(\tau' \in Black\right) s^{\tau'}_{m} \, \hat{p}^{\tau'}_{\ell,m}}{\sum_{\tau} s^{\tau}_{m} \, \hat{p}^{\tau}_{\ell,m}}$$

This is the same as equation (8) with the exception that $\hat{p}^{\tau}_{\ell,m}$ is the observed, not predicted, probability that household type τ chooses location ℓ in metro m. $S^{b}_{\ell,m}$ is an endogenous variable because $\hat{p}^{\tau}_{\ell,m}$ depends on unobserved amenities, $A^{\tau}_{\ell,m}$. When the predicted Black share $Z^{b}_{\ell,m}$ is constructed using equation (8), the observed probability $\hat{p}^{\tau}_{\ell,m}$ is replaced with a predicted probability $\hat{\hat{p}}^{\tau}_{\ell,m}$ that only depends on topographic data. When we replace the observed probabilities with the predicted probabilities to construct the instrument, conditional on topography all of the variation in predicted Black shares $Z^{b}_{\ell,m}$ will be attributable to variation in household type shares across metro areas.

To see this, we now work through a simple illustrative example with two metros where the metros have the identical topographic map but different metro-wide shares of household types. Suppose:

a. There are only two types of households in the economy, Black and White. The share of Black households in the first metro area "A" is 5 percent and the share of Black households in the second metro area "B" is 40 percent.
b. Topography in each of metros A and B can be characterized as 50 percent "relatively flat" and 50 percent "relatively elevated."

2. A regression of location choice data with national data like equation (7) predicts that Black households choose to live in the relatively flat part of a metro area 65 percent of the time and White households choose to live in the relatively flat part of a metro area 45 percent of the time.

In this fictitious example, the table below shows how the predicted share of Black households varies from relatively flat to relatively elevated parts of metros A and B

	Black Share	
Location	Metro A	Metro B
Relatively Flat	7.1%	49.1%
Relatively Elevated	3.2%	29.8%

In both metro A and B, the share of Black households increases from the relatively elevated to the relatively flat parts of the metro. However, the predicted level of the Black share in all parts of the metro and – more importantly – the predicted change in the Black share from the relatively elevated to the relatively flat part of the metro are much larger in metro B than in metro A. The shares in this table are computed according to equation (8) as follows

$$
\begin{array}{llclcl}
 & \text{A} & & & \text{B} & \\
\text{flat} & 0.071 & = & \dfrac{\mathbf{0.05} \times 0.65}{\mathbf{0.05} \times 0.65 \ + \ \mathbf{0.95} \times 0.45} & 0.491 & = \dfrac{\mathbf{0.40} \times 0.65}{\mathbf{0.40} \times 0.65 \ + \ \mathbf{0.60} \times 0.45} \\
\text{elevated} & 0.032 & = & \dfrac{\mathbf{0.05} \times 0.35}{\mathbf{0.05} \times 0.35 \ + \ \mathbf{0.95} \times 0.55} & 0.298 & = \dfrac{\mathbf{0.40} \times 0.35}{\mathbf{0.40} \times 0.35 \ + \ \mathbf{0.60} \times 0.55}
\end{array}
$$

The boldface text highlights that the only variable generating differences in the predicted Black shares between metros A and B for a given topography is the metro-area share of the population that is Black. As equations (9) and (10) show, when we estimate preference parameters using 2SLS we include metro fixed effects and control for the direct impact of topography on location choice, so the variation we emphasize in this example – variation in the shares of Black and White households across metro areas – is driving identification.

2.2.3 Discussion of the Exclusion Restrictions

The recent literature on the econometrics of shift-share instrumental variables has emphasized that either the exogeneity of shares (Goldsmith-Pinkham, Sorkin, and Swift, 2020) or the exogeneity of shifts (Borusyak, Hull, and Jaravel, 2022), but not necessarily both, is required for the shift-share IVs resulting from the aggregation of the shift-share interactions to satisfy instrument exogeneity. We proceed under an assumption of share exogeneity, which can be stated formally as follows:

A1) Share exogeneity:

$$\text{across m:} \qquad \ddot{A}^{\tau}_{\ell,m} \;\perp\; s^{\tau'}_{m} \Big| \ddot{TOP}_{\ell,m} \quad \forall \tau, \tau' \tag{11}$$

where $\ddot{X}$ means that variable X has had metro-area fixed effects removed. In words, this equation specifies the following: Consider a given topography and a given type of household τ. Across metro areas, the amenity value (for that τ) at locations with that specific topography cannot be correlated with the realized metro-level population shares of any type of household. This lack of correlation must hold for all types τ and all possible topographies.

Under the assumption of share exogeneity, the shift variables (national-average predicted neighborhood choice probabilities based only on topography) can be thought of as weights that capture the idea that city-wide metro shares of particular types contribute more strongly or less strongly to predicted neighborhood-level racial makeup depending on which topographic features are present in the neighborhood. Because we control for the direct effect of topography on location choices, identification in our framework is driven only by comparisons of tracts with similar topographic configurations that have different predicted racial compositions due to the metro-wide type shares in the metro where they are located.

In seminars, we have been asked to more intuitively explain what is required for identification. To do this, we revisit equation (9), rewritten here for convenience:

$$\hat{d}^{\tau}_{\ell,m} \;=\; \theta^{\tau}_{m} \;+\; a^{\tau}_{1} S^{b}_{\ell,m} \;+\; g^{\tau}\left(TOP_{\ell,m}\right) \;+\; \upsilon^{\tau}_{\ell,m} \tag{12}$$

Note that for consistent estimation of a_1, it does not matter if topography is endogenous and correlated with $\upsilon^{\tau}_{\ell,m}$ since equation (12) explicitly controls for topography. Now apply the Frisch-Waugh-Lovell theorem: regress each of $\hat{d}^{\tau}_{\ell,m}$ and $S^{b}_{\ell,m}$ on metro fixed effects and $g^{\tau}\left(TOP_{\ell,m}\right)$. Denote the residuals from this regression as $\ddot{\hat{d}}^{\tau}_{\ell,m}$ and $\ddot{S}^{b}_{\ell,m}$ and rewrite equation (12) as

$$\ddot{\hat{d}}^{\tau}_{\ell,m} \;=\; a^{\tau}_{1} \ddot{S}^{b}_{\ell,m} \;+\; \ddot{\upsilon}^{\tau}_{\ell,m}$$

The error term in this equation, $\ddot{\upsilon}^{\tau}_{\ell,m}$, is uncorrelated with metro fixed effects and uncorrelated with topography. Thus, a valid instrument only needs to be uncorrelated with the component of amenities that is uncorrelated with topography. Since the variation in type shares across metros generates the variation in the instrument we construct, this explains our required assumption A1 that across-metro variation in type shares needs to be uncorrelated with the component of amenities that is itself uncorrelated with topography.

2.2.4 Discussion of the Great Migration of 1910-1970

For share exogeneity to be violated, household types would have needed to systematically choose metros in which to locate based on the component of amenities (uncorrelated with topography) in neighborhoods in that metro with the topographic features commonly chosen by their household type. To be clear, share exogeneity is *not* violated if household types chose metros in which to locate based on metro-wide factors or amenities common to all topographies in the metro.

The most prominent historical process of voluntary metro-level location choices affecting modern metro-level racial composition was the Great Migration of 1910-1970 in which roughly 6 million moved from the rural South to Northern metro areas. Therefore, understanding the factors that most strongly influenced households' choices of a Northeren metro during this period provides a direct check of the plausibility of our identifying assumption. Share exogeneity would likely be violated if, during this period, Black households moved to metro areas based on the relative amenities (not directly driven by topography) of the topographies in each metro in which they were most likely to locate, as compared to the overall desirability of the metro area. If these considerations played a meaningful role in determining the metros to which Black households migrated, we would expect a positive correlation between $s_m^{\tau'}$ for Black household types and $\ddot{A}^{\tau}_{\ell,m}$ at the topographies where those households are most likely to live, violating share exogeneity – assumption A1 as written in equation (11).

Stuart and Taylor (2021) provides a summary of the literature on factors that influenced location decisions during the Great Migration. These factors include: 1) Labor demand factors such as Word War I, which increased manufacturing employment in Northern cities and decreased labor supply from European immigrants; 2) Southern push factors such as the destruction of cotton crops by the boll weevil, labor market discrimination, racial violence, and Jim Crow laws; 3) railroad networks which channeled migration from certain regions in the South to specific Northern cities, and; 4) social networks of family, friends, and church members that had already migrated to the North. None of these factors are related to the amenities of the neighborhoods at the within-metro relative topographies in which Black households tend to live. For these reasons, we think it unlikely that the Great Migration led to migration flows that violate share exogeneity.

2.3 Monte Carlo Simulations

We now perform a set of Monte Carlo experiments using this simplified version of our model to demonstrate the ability of our proposed estimator to recover accurate estimates of racial preferences for multiple household types in a setting with a number of metros and neighborhoods similar to that of our actual sample. In the experiments, we repeatedly simulate decisions of 4 types of households in 200 metro areas, each with 100 possible locations. Each time we simulate the model using procedures described next, we generate a data set containing the information listed in section 2.2.1 that is required for estimation: Type shares, neighborhood choice probabilities, topographic data, and rental prices. We then estimate model parameters using that data via the IV procedure described in that section. From each simulation, we store the estimate of the model parameter of interest, a_1^τ, and then evaluate the mean and standard deviation of this estimate across simulations to understand the size of possible bias. In the rest of this section, we list the details of the simulation procedure and results.

2.3.1 Drawing Metro-Wide Type Shares and Local Amenities

We specify that type 1 and 2 households are Black and type 3 and 4 households are White. To determine the simulated share of type τ in metro m, s_m^τ, we first compute the variable $\tilde{s}_m^\tau$ as

$$\tilde{s}_m^\tau \quad = \quad \ln \mu^\tau + e_m^\tau \tag{13}$$

where e_m^τ is a random draw from a Normal with mean 0 and standard deviation σ_{em}^τ. In all simulations, we set $\mu^\tau = 0.25$ and $\sigma_{em}^\tau = 1.0$ for all household types. We then set simulated type shares in each metro as as

$$s_m^\tau \quad = \quad \frac{\exp\{\tilde{s}_m^\tau\}}{\sum\limits_{\tau'=1}^{4} \exp\{\tilde{s}_m^{\tau'}\}} \tag{14}$$

The total population in every metro is assumed to be 1.0, implying the share of each type is also the population of that type.

Amenities for each location ℓ in each metro m for each type τ are assumed to be a linear function of a topography variable $TOP_{\ell,m}$ and 4 other common factors, $f_{n,\ell,m}$ for $n = 1, \ldots, 4$, as follows

$$\mathcal{A}_{\ell,m}^\tau \quad = \quad \sum_{n=1}^{4} \gamma_n^\tau f_{n,\ell,m} \; + \; \gamma_{top}^\tau TOP_{\ell,m}$$

Each of the four common factors and the topography variable can vary across locations within and across metros, but for any given location in any given metro these variables do not vary by type. The parameters $\gamma_n^\tau, n = 1, \ldots, 4$ and γ_{top}^τ can vary across types of households but for any given type of household they do not vary across locations and metros. We chose four factors to allow us to consider a scenario where each type's valuations of amenities (other than topography) are independent. In the simulations, we draw $f_{n,\ell,m}$ and $TOP_{\ell,m}$ iid from $N(0,1)$.

The parameters we use in our baseline simulations for each type are listed in Table 1. Note that we set $\gamma_{top}^\tau > 0$ for all types, implying all types value the topography variable, but by varying degrees. Note that for a_r^τ, we have in mind that Types 1 and 3 have lower income and higher budget shares on rent than Types 2 and 4. Additionally, Types 1 and 3 are less willing to pay for high values of the topography variable, the component of location-specific amenities that is common to all types.

Table 1: Baseline Set of Parameters for Monte Carlo

Type	Race	a_r^τ	a_1^τ	γ_1^τ	γ_2^τ	γ_3^τ	γ_4^τ	γ_{top}^τ
1	Black	0.5	0.5	1	0	0	0	0.25
2	Black	0.3	0.5	0	1	0	0	0.75
3	White	0.4	-0.5	0	0	1	0	0.50
4	White	0.2	-0.5	0	0	0	1	1.00

Notes: In the baseline parameterization, we assume housing in every location in every metro area is inelastically supplied. Referring to equation (13), we set $\mu^\tau = 0.25$ and $\sigma_{em}^\tau = 1.0$ for all types.

2.3.2 Finding Rental Prices and Racial Shares

For a given metro area m, we need to find market-clearing rental prices $r_{\ell,m}$ and, given these prices, the share of Black households $S_{\ell,m}^b$ in each of the locations ℓ in the metro. We employ the following algorithm to find $r_{\ell,m}$ and $S_{\ell,m}^b$ in every ℓ in every m:

a. We initialize $r_{\ell,m}$ and $S_{\ell,m}^b$ by setting $r_{\ell,m} = 1$ and $S_{\ell,m}^b$ equal to the metro-m wide share of black households $(s_m^1 + s_m^2)$ for every metro. Label these initial values as $\hat{r}_{\ell,m}$ and $\hat{S}_{\ell,m}^b$.

b. Given the simulated realized values of $\mathcal{A}_{\ell,m}^\tau$ and the values $\hat{r}_{\ell,m}$ and $\hat{S}_{\ell,m}^b$, we compute

$$\hat{\delta}_{\ell,m}^\tau \;=\; -a_r^\tau \log \hat{r}_{\ell,m} \;+\; a_1^\tau \hat{S}_{\ell,m}^b \;+\; \mathcal{A}_{\ell,m}^\tau$$

for every location ℓ and every type τ.

c. After steps a and b have been completed for every ℓ and k in every metro m for every household type τ, we compute for every location ℓ in metro m and type τ

$$\hat{p}^{\tau}_{\ell,m} \quad = \quad \frac{\exp\left(\hat{\delta}^{\tau}_{\ell,m}\right)}{\sum_{k}\exp\left(\hat{\delta}^{\tau}_{k,m}\right)}$$

where $\hat{p}^{\tau}_{\ell,m}$ is the probability that type τ chooses location ℓ in metro m.

d. Give the distribution of all location choices for all types in all metros as determined in step c, we compute the simulated population in each location ℓ in metro m, $\widetilde{pop}_{\ell,m}$, and the simulated Black share of the population, $\widetilde{S}^{b}_{\ell,m}$ as

$$\begin{aligned}
\widetilde{pop}_{\ell,m} &= \sum_{\tau=1}^{4}\hat{p}^{\tau}_{\ell,m}s^{\tau}_{m} \\
\widetilde{S}^{b}_{\ell,m} &= \left(\hat{p}^{1}_{\ell,m}s^{1}_{m}+\hat{p}^{2}_{\ell,m}s^{2}_{m}\right)/\widetilde{pop}_{\ell,m}
\end{aligned}$$

e. For each ℓ, we update $\hat{S}^{b}_{\ell,m}$ by setting it equal to $\widetilde{S}^{b}_{\ell,m}$

f. When housing is inelastically supplied, we assume (i) every location ℓ has 0.01 units of housing, equal to the total population divided by 100 locations, and (ii) each household demands 1 unit of housing. We compute

$$\log\widetilde{r}_{\ell,m} \quad = \quad \log\hat{r}_{\ell,m} \;+\; \log\widetilde{pop}_{\ell,m} \;-\; \log 0.01$$

and then we set the updated value of $\hat{r}_{\ell,m}$ equal to $\widetilde{r}_{\ell,m}$.

g. We repeat steps b through f for all locations until expectations on the share of Black households converges to simulated Black shares in every location ($\widetilde{S}^{b}_{\ell,m} = \hat{S}^{b}_{\ell,m}$), and, $\widetilde{pop}_{\ell,m} = 0.01$ in every location.

2.3.3 Other Features and Assumptions

We simulate the model under 5 different sets of assumptions, listed below. Unless otherwise stated, each environment uses the baseline set of parameters.

1. *Baseline.* Baseline set of parameters.
2. *Elastic Supply.* In these simulations, we set and hold $r_{\ell,m} = 1.0$ everywhere; convergence (steps a-g listed above) requires only that expectations on the share of Black households is equal to simulated Black shares in every location, $\widetilde{S}^{b}_{\ell,m} = \hat{S}^{b}_{\ell,m}$.
3. *Low Variance.* $\sigma^{\tau}_{em} = 0.3$. This simulation helps show the relationship of the variance of estimates of a^{τ}_{1} to the amount of across-metro variation in household type shares.
4. *Correlated Amenities.* $\gamma^{\tau}_{\tau'} = 0.5$ for $\tau' \neq \tau$. This simulation checks the sensitivity of

estimates to assumptions about how much each household type also cares about the favorite amenity of the other household types by increasing $\gamma^{\tau}_{\tau'}$ from 0 in the baseline to 0.5.

5. *Imperfect Factor Measurement.* Location-specific amenities are determined as

$$\begin{aligned} \mathcal{A}^{\tau}_{\ell,m} &= \sum_{n=1}^{4} \gamma^{\tau}_{n} f_{n,\ell,m} + \gamma^{\tau}_{top} f_{5,\ell,m} \\ f_{5,\ell,m} &= TOP_{\ell,m} + f_{6,\ell,m} \end{aligned}$$

where the topography variable is observed by the econometrician (as in the baseline) but $f_{6,\ell,m}$, also drawn from $N(0,1)$, is not observed. What we have in mind is that the topography that households value in amenities, in this case $f_{5,\ell,m}$, is only imperfectly proxied by the topography variable that an econometrician observes, $TOP_{\ell,m}$. This simulation checks that results are robust to imperfect observation of topography.

For each of the 5 environments described above, we generate 1,000 data sets and evaluate the properties of two different estimators for a^{τ}_{1} in equation (9): OLS and our IV. For a final set of details, note that $g(TOP_{\ell,m})$ and $g_1(TOP_{\ell,m})$ are both 4th order polynomials and we use exact market shares as predicted by the model as our data for $\hat{p}^{\tau}_{\ell,m}$. Finally, to be explicit, we assume an econometrician observes $TOP_{\ell,m}$ for all ℓ in every m but $f_{n,\ell,m}$ for $n = 1, \ldots, 4$ and $A^{\tau}_{\ell,m}$ are never observed.

2.3.4 Results

The results of this exercise are shown in Table 2. The eight columns of the table report the average and standard deviation of estimates of $a^{\tau}_{1}, \tau = 1, \ldots, 4$ across all 1,000 data sets. The first column of each pair shows estimates when a^{τ}_{1} is estimated using OLS and the second shows estimates using IV. Recall the actual value of a^{τ}_{1} is 0.5 for Types 1 and 2 and -0.5 for Types 3 and 4, as shown in Table 1.

Table 2 clearly shows that the OLS estimates are biased. There is no mean OLS estimate of any parameter that can be considered "close" in any meaningful sense to the corresponding preference parameter. This highlights the challenge of identification. Household sorting based on type-specific valuations of amenities leads to correlation of the amenities in the location and the share of Black households in that location. Even though OLS controls for topography, a common component of amenities, the omission of the other factors $f_{1,\ell,m}, \ldots, f_{4,\ell,m}$ that determine amenities (along with weights in the computation of amenities $\gamma^{\tau}_{1}, \ldots, \gamma^{\tau}_{4}$ that vary by type) is sufficient to cause OLS to be severely biased.

In contrast, for each parameter we consider in all 5 simulation scenarios, the IV estimator has a mean (across the 1,000 simulations) very close to 0.5 for $\tau = 1, 2$ and -0.5 for $\tau = 3, 4$. Even though the environment changes in each of the 5 scenarios, what is common across scenarios is that the distribution of type shares in each metro area is drawn independently of the realization of all factors and topography comprising amenities in any given metro area. As shown by equation (11), this is a sufficient condition for our IV approach to be valid.

Table 2: Monte Carlo Results

		$\tau = 1$		$\tau = 2$		$\tau = 3$		$\tau = 4$	
		$a_1^\tau = 0.5$		$a_1^\tau = 0.5$		$a_1^\tau = -0.5$		$a_1^\tau = -0.5$	
Sim	Name	OLS	IV	OLS	IV	OLS	IV	OLS	IV
1	Baseline	1.785	0.494	2.054	0.495	-2.367	-0.499	-2.311	-0.500
		(0.056)	(0.146)	(0.062)	(0.133)	(0.064)	(0.144)	(0.071)	(0.144)
2	Elastic Supply	2.340	0.484	2.133	0.490	-2.430	-0.493	-2.044	-0.498
		(0.057)	(0.252)	(0.058)	(0.235)	(0.055)	(0.252)	(0.059)	(0.251)
3	Low Variance	1.881	0.485	2.118	0.485	-2.505	-0.491	-2.371	-0.490
		(0.032)	(0.349)	(0.032)	(0.319)	(0.031)	(0.350)	(0.034)	(0.345)
4	Correlated Amenities	-1.429	0.493	-1.216	0.494	-4.309	-0.504	-4.165	-0.505
		(0.126)	(0.172)	(0.158)	(0.165)	(0.119)	(0.169)	(0.150)	(0.171)
5	Imperfect Factor Msmt	1.368	0.495	1.152	0.500	-2.741	-0.495	-3.098	-0.493
		(0.049)	(0.171)	(0.093)	(0.190)	(0.056)	(0.184)	(0.099)	(0.227)

3 The Actual Dynamic Location-Choice Model

We model the system of demand for neighborhoods by considering the decision problem of a particular household head deciding where the household should live. As in Kennan and Walker (2011) Bayer, McMillan, Murphy, and Timmins (2015), and Davis, Gregory, Hartley, and Tan (2021) we model location choices in a dynamic discrete choice setting. We assume each household i takes its metro area m as given. Each year, the household can choose to live in one of J_m locations in the metro. When we estimate this model, J_m will vary with the metro.

Denote j as the current location of household i in the metro and τ as that household's type. We write the value to the household, $V^{\tau}_{m,t}\left(\ell \mid j, \epsilon^{\tau}_{i,\ell,m,t}\right)$, of choosing to live in location ℓ in metro m in year t given a current location of j in the metro and current value of a shock $\epsilon^{\tau}_{i,\ell,m,t}$ (to be explained later) as

$$V^{\tau}_{m,t}\left(\ell \mid j, \epsilon^{\tau}_{i,\ell,m,t}\right) \;\;=\;\; u^{\tau}_{m,t}\left(\ell \mid j, \epsilon^{\tau}_{i,\ell,m,t}\right) + \beta \sum_{\tau'} \varphi^{\tau,\tau'} E_t \left[V^{\tau'}_{m,t+1}\left(\ell\right)\right]$$

In the above equation $u^{\tau}_{m,t}\left(\ell \mid j, \epsilon^{\tau}_{i,\ell,m,t}\right)$ is the flow utility in year t to household i of choosing to live in location ℓ in metro m given a current location of j in the metro and current value of a shock $\epsilon^{\tau}_{i,\ell,m,t}$; β is the discount factor on future expected utility; $\varphi^{\tau,\tau'}$ is the probability that the household becomes type τ' next year given it is type τ this year; and $E_t\left[V^{\tau'}_{m,t+1}\left(\ell\right)\right]$ is the expected value (expectation taken as of year t) in year $t+1$ of a type τ' household of having chosen to live in neighborhood ℓ in metro m in year t. The t subscripts explicitly allow that flow utility and expectations may change over time.

Flow utility depends on neighborhood racial composition, similar to assumptions made in Caetano and Maheshri (2021) and Almagro and Dominguez-Iino (2022).[8] We specify $u^{\tau}_{m,t}\left(\ell \mid j, \epsilon^{\tau}_{i,\ell,m,t}\right)$ as

$$u^{\tau}_{m,t}\left(\ell \mid j, \epsilon^{\tau}_{i,\ell,m,t}\right) \;\;=\;\; \delta^{\tau}_{\ell,m,t} \;-\; \kappa^{\tau} \cdot 1_{\ell \neq j} \;+\; \epsilon^{\tau}_{i,\ell,m,t}$$

$\delta^{\tau}_{\ell,m,t}$ is the deterministic portion of flow utility a type τ household receives in year t from living in neighborhood ℓ in metro m; this does not vary across type τ households. κ^{τ} are fixed costs a household of type τ must pay when it moves to a different neighborhood in the metro i.e. when $\ell \neq j$; $1_{\ell \neq j}$ is an indicator function that is equal to 1 if location $\ell \neq j$ in metro m and 0 otherwise; and $\epsilon^{\tau}_{i,\ell,m,t}$ is a random shock that is known at the time of the location choice. $\epsilon^{\tau}_{i,\ell,m,t}$ is assumed to be iid across locations, time and people. $\epsilon^{\tau}_{i,\ell,m,t}$ induces otherwise identical households living at the same location at the same time to optimally choose different future locations.

We assume $\delta^{\tau}_{\ell,m,t}$ is comprised of disutility from log rental prices ($\log r_{\ell,m,t}$), a quadratic function of the share of neighborhood ℓ that is Black ($S^{b}_{\ell,m,t}$) and is Hispanic ($S^{h}_{\ell,m,t}$), and

[8] The utility function in Almagro and Dominguez-Iino (2022) does not depend on race but does depend on neighborhood consumption amenities which are endogenously determined.

amenities of that neighborhood that may vary by household type, $\mathcal{A}^{\tau}_{\ell,m,t}$.

$$\begin{aligned}\delta^{\tau}_{\ell,m,t} \;=\;& -a^{\tau}_{r}\log r_{\ell,m,t} && \text{rents}\\ & +a^{\tau}_{1}S^{b}_{\ell,m,t}+a^{\tau}_{2}\left(S^{b}_{\ell,m,t}\right)^{2}+a^{\tau}_{3}S^{h}_{\ell,m,t}+a^{\tau}_{4}\left(S^{h}_{\ell,m,t}\right)^{2}+a^{\tau}_{5}S^{b}_{\ell,m,t}S^{h}_{\ell,m,t} && \text{demographics}\\ & +\mathcal{A}^{\tau}_{\ell,m,t} && \text{amenities}\end{aligned} \tag{15}$$

We do not impose a linear specification in racial shares because we do not want to impose that the marginal utility of a change in a racial share is constant with respect to the level of that racial share. A quadratic functional form is a parsimonious specification that allows for the possibility that households may like some diversity; it also allows, depending on parameters, that households may not like any diversity.

Denote $\epsilon^{\tau}_{1,m,t}$ as the shock associated with location 1 in period t, $\epsilon^{\tau}_{2,m,t}$ as the shock with location 2, and so on. In each period after the vector of ϵ are revealed (one for each location), households choose the location that yields the maximal value

$$V^{\tau}_{m,t}\left(j \mid \epsilon^{\tau}_{1,m,t},\epsilon^{\tau}_{2,m,t},\ldots,\epsilon^{\tau}_{J_m,m,t}\right) \;=\; \max_{\ell\in 1,\ldots,J_m} V^{\tau}_{m,t}\left(\ell \mid j,\epsilon^{\tau}_{i,\ell,m,t}\right) \tag{16}$$

4 Estimation and Data

To match model to data, we assume that a location (neighborhood) is a census tract. We use a 2-step procedure like Berry, Levinsohn, and Pakes (1995) to estimate our model of demand for locations. In the first step, we use GMM to estimate the vector of $\delta^{\tau}_{\ell,m,t}$ and the moving cost κ^{τ} for each τ. This is similar to the procedure of Neilson (2017), who uses GMM to estimate a similar first stage in a model of school choice. In the second step, we use an IV procedure to understand how exogenous changes in racial shares impact $\delta^{\tau}_{\ell,m,t}$ for each τ.

4.1 Step 1: GMM to Estimate Demand for Locations

In the first step, we use the approach of Hotz and Miller (1993) employed by Bishop (2012) and Davis, Gregory, Hartley, and Tan (2021) to set up estimating equations for $\delta^{\tau}_{\ell,m,t}$ and κ^{τ}. This approach does not require that we solve for the value functions. Instead, as we show in appendix A, the log probabilities that choices are observed are simple functions of $\delta^{\tau}_{\ell,m,t}$, κ^{τ}, β and of observed choice probabilities. Note that due to data limitations we discuss later, we combine data across multiple years when estimating probabilities and preference parameters. For this reason, going forward we remove time subscripts from value functions, expectations and elements of utility.

Define Θ_1^τ as the full vector of parameters to estimate in step 1 for type τ

$$\Theta_1^\tau = \left\{ \kappa^\tau, \left\{\delta_{\ell,m=1}^\tau\right\}_{\ell=1}^{J_1}, \left\{\delta_{\ell,m=2}^\tau\right\}_{\ell=1}^{J_2}, \ldots, \left\{\delta_{\ell,m=M}^\tau\right\}_{\ell=1}^{J_M} \right\} \tag{17}$$

where $\delta_{\ell,m}^\tau$ is the value of δ for type τ in tract ℓ in metro m, assumed fixed over the years in our estimation sample, and M is the number of metros in the sample.

The first moment we target for each household type is the unconditional probability of not moving. Define the distance between the model predicted non-moving rate and the data as

$$\begin{aligned} & G_1^\tau\left(\Theta_1^\tau\right) \\ & = \sum_{m=1}^{M}\sum_{j'=1}^{J_m} \underbrace{\hat{P}_m^\tau\left(j=j'\right)}_{\text{data}} \underbrace{\hat{P}_m^\tau\left(\ell=j' \mid j=j'\right)}_{\text{data}} - \sum_{m=1}^{M}\sum_{j'=1}^{J_m} \underbrace{\hat{P}_m^\tau\left(j=j'\right)}_{\text{data}} \underbrace{P_m^\tau\left(\ell=j' \mid j=j';\Theta_1^\tau\right)}_{\text{model}} \end{aligned} \tag{18}$$

In this equation, j is the location at the start of the period and ℓ is the location at the end of the period. j' indexes locations that are in metro m and there are J_m of these locations. In this equation and the next, any variable with a "hat" is computed directly from the data. $\hat{P}_m^\tau(j=j')$ is the probability that a type τ household starts a period in location j' in metro m and $\hat{P}_m^\tau(\ell=j' \mid j=j')$ is the probability that a type τ household that starts a period in location j' chooses to remain in location j'. The conditional probability $P_m^\tau(\ell \mid j;\Theta_1^\tau)$ for any ℓ and j is determined by the model for a given Θ_1^τ.

The remaining $\sum\limits_{m=1}^{M}[J_m-1]$ moments for each type are that the model matches the probability of choosing any given location in each metro. There are J_m-1 moments in each metro because the probability of choosing a location must sum to 1, and (as mentioned) households are assumed to not move outside of their metro. For any given metro m, we can write the distance for these J_m-1 moments as

$$\text{for } \ell=2,\ldots,J_m \quad G_{\ell,m}^\tau\left(\Theta_1^\tau\right) = \sum_{j=1}^{J_m} \underbrace{\hat{P}_m^\tau(j)}_{\text{data}} \underbrace{\hat{P}_m^\tau(\ell \mid j)}_{\text{data}} - \sum_{j=1}^{J_m} \underbrace{\hat{P}_m^\tau(j)}_{\text{data}} \underbrace{P_m^\tau(\ell \mid j;\Theta_1^\tau)}_{\text{model}} \tag{19}$$

We normalize $\delta_{1,m}^\tau = 0$ in each metro, which is allowable because utility is relative and adding a constant to each δ^τ in the choice set will not affect the probability of any choice.

For each type τ, we find the vector of parameters to minimize the sum of squared errors of the moments

$$\hat{\Theta}_1^\tau = \underset{\Theta_1^\tau}{\operatorname{argmin}} \left\{ \left[G_1^\tau\left(\Theta_1^\tau\right)\right]^2 + \sum_{m=1}^{M} \sum_{\ell=2}^{J_m} \left[G_{\ell,m}^\tau\left(\Theta_1^\tau\right)\right]^2 \right\}$$

The model is exactly identified, so at $\Theta_1^\tau = \hat{\Theta}_1^\tau$ the term in braces will be zero. For each type, there are $1 + \sum\limits_m (J_m - 1)$ moments and the same number of parameters.

4.2 Step 2: IV to Estimate Impact of Demographics on Demand

Once we have estimates of $\delta_{\ell,m}^\tau$ from the 1st stage, we wish to uncover the parameters a_r^τ and $a_1^\tau, \ldots, a_5^\tau$ from equation (15). We start by taking a value for the impact of rental prices on flow utility, a_r^τ, from Davis, Gregory, Hartley, and Tan (2021). Define $\hat{\delta}_{\ell,m}^\tau$ as the estimated value of $\delta_{\ell,m}^\tau$ from the first stage. Then we wish to estimate $a_1^\tau, \ldots, a_5^\tau$ in the following specification that is the same as equation (15), but with time subscripts removed:

$$\hat{\delta}_{\ell,m}^\tau + a_r^\tau \log r_{\ell,m} = a_1^\tau S_{\ell,m}^b + a_2^\tau \left(S_{\ell,m}^b\right)^2 + a_3^\tau S_{\ell,m}^h + a_4^\tau \left(S_{\ell,m}^h\right)^2 + a_5^\tau S_{\ell,m}^b S_{\ell,m}^h + A_{\ell,m}^\tau \quad (20)$$

In the equation above, $\log r_{\ell,m}$ is an estimate of the log rental price for a standardized housing unit in neighborhood (census tract) ℓ of metro m that we compute from data from the 2007-2011 American Community Survey tract-level tabulations.[9] Also note, consistent with our practice in section 2, the switch of notation for amenities from $\mathcal{A}$ in equation (15) to A in the above equation.

Define the vector of parameters that we estimate for each type in this second step as Θ_2^τ,

$$\Theta_2^\tau = \{ a_1^\tau, a_2^\tau, \ldots, a_5^\tau \}$$

Since racial shares are likely to be correlated with unobserved amenities, we use an instrumental variables approach to estimate Θ_2^τ. Our approach is nearly identical to what we do in our that described in the Monte Carlo section. To generate instruments for $S_{\ell,m}^b$ and $S_{\ell,m}^h$,

[9] Define $\bar{r}_{\ell,m}$ as the median rent of renting households for census tract ℓ in metro m and $\mathbf{X}_{\ell,\mathbf{m}}$ as a vector of tract-level housing characteristics. To compute $\log r_{\ell,m}$, we first run the regression

$$\log \bar{r}_{\ell,m} = [\mathbf{\Gamma}]' \mathbf{X}_{\ell,m} + e_{\ell,m}^r$$

where $\mathbf{\Gamma}$ is a vector of coefficients and $e_{\ell,m}^r$ is the error in this equation. Defining $\hat{e}_{\ell,m}^r$ as the estimate of the residual from this regression, we set $\log r_{\ell,m} = \left[\hat{\mathbf{\Gamma}}\right]' \bar{\mathbf{X}} + \hat{e}_{\ell,m}^r$ where $\hat{\mathbf{\Gamma}}$ is the vector of coefficient estimates of $\mathbf{\Gamma}$ and $\bar{\mathbf{X}}$ is the average value (element by element) of $\mathbf{X}_{\ell,\mathbf{m}}$.

for each type of household we pool all data across locations and metros and estimate

$$\log \hat{p}^{\tau}_{\ell,m} \;=\; \alpha^{\tau}_{m} \;+\; [\mathbf{b}^{\tau}]' \cdot \mathbf{TOP}_{\ell,m} \;+\; \nu^{\tau}_{\ell,m} \tag{21}$$

where $\hat{p}^{\tau}_{\ell,m}$ is the estimated market share (within-metro choice probability) for location ℓ in metro m for type τ households, α^{τ}_{m} is a metro-area fixed effect that can vary by τ, and $\nu^{\tau}_{\ell,m}$ is an error. In comparison to the prediction equation (7) of the simple model, in equation (21) $\mathbf{TOP}_{\ell,m}$ is a vector of topographic variables for each location ℓ in metro m and $\mathbf{b}^{\tau}$ is a vector of coefficients that can vary by τ.

Once equation (21) is estimated for all types, we use it to predict the probability each type lives in any location in any metro area. Call this predicted probability as $\hat{\hat{p}}^{\tau}_{\ell,m}$. Given these predicted choice probabilities for all types in all locations in all metros, we create a predicted Black share $Z^{b}_{\ell,m}$ and a predicted Hispanic share $Z^{h}_{\ell,m}$ in each ℓ and m as as

$$Z^{b}_{\ell,m} \;=\; \frac{\sum\limits_{\tau'} \mathcal{I}\left(\tau' \in Black\right) s^{\tau'}_{m}\, \hat{\hat{p}}^{\tau'}_{\ell,m}}{\sum\limits_{\tau} s^{\tau}_{m}\, \hat{\hat{p}}^{\tau}_{\ell,m}} \tag{22}$$

$$Z^{h}_{\ell,m} \;=\; \frac{\sum\limits_{\tau'} \mathcal{I}\left(\tau' \in Hispanic\right) s^{\tau'}_{m}\, \hat{\hat{p}}^{\tau'}_{\ell,m}}{\sum\limits_{\tau} s^{\tau}_{m}\, \hat{\hat{p}}^{\tau}_{\ell,m}} \tag{23}$$

In the above equations τ and τ' are indexes for household type; s^{τ}_{m} is the share of the metro population that is accounted for by type τ households; $\mathcal{I}\left(\tau' \in Black\right)$ is an indicator function that is equal to 1 if type τ' households are Black, 0 otherwise; and $\mathcal{I}\left(\tau' \in Hispanic\right)$ is an indicator function that is equal to 1 if type τ' households are Hispanic, 0 otherwise. When the population of the metro area is equal to 1.0, the numerators of (22) and (23) are the predicted total population of Black and Hispanic households, respectively, living in tract ℓ in metro m and the denominator of both equations is the total population of households living in tract ℓ in metro m. In both equations, the predictions are based only on the topography variables, equation (21). We use equation (21) to predict where everyone will live based only on topography and then, given this prediction, use metro-level household-type shares to calculate the predicted share of each location that is Black and Hispanic.

Our five instruments are then $Z^{b}_{\ell,m}$, $Z^{h}_{\ell,m}$, $\left(Z^{b}_{\ell,m}\right)^2$, $\left(Z^{h}_{\ell,m}\right)^2$, and $\left(Z^{b}_{\ell,m} Z^{h}_{\ell,m}\right)$. With these instruments in hand, we estimate Θ^{τ}_{2} of equation (20) using 2SLS for each type, exactly analogous to the procedure we use for the Monte Carlo of the simple model described earlier. In both the first and second stages, we include metro-area fixed effects and a set of topographic variables as controls. We discuss the topographic data later in this section.

4.3 Data

4.3.1 FRBNY Consumer Credit Panel / Equifax

We estimate the model using panel data from the FRBNY Consumer Credit Panel / Equifax data set (CCP). The panel is comprised of a 5% random sample of U.S. adults with a social security number, conditional on having an active credit file, and any individuals residing in the same household as an individual from that initial 5% sample.[10] For years 1999 to 2019, the database provides a quarterly record of variables related to debt: Mortgage and consumer loan balances, payments and delinquencies and a few other variables we discuss later. The data does not contain information on basic demographics like race, education, or number of children. The data also includes the Equifax Risk Score^{TM} which provides some information on the financial wherewithal of the household as demonstrated in Board of Governors of the Federal Reserve System (2007).

Most important for our application, the panel data includes in each period the current census block (and therefore census tract) of residence.[11] To match the annual frequency of our location choice model, we use location data from the first quarter of each calendar year. In each year, we only include people living in metro areas – if, for example, a household moves from an eligible metro area to a rural area, that household-year observation is not included in the estimation sample. To keep estimation computationally feasible, we assume each household can only move within its metropolitan division ("metro"). If a household moves to a different metro, the household-year observation of the move is not included in the estimation sample, but the years before and after the across-metro move are included.

The panel is not balanced, as some individuals' credit records first become active after 1999. We restrict the sample to households living in one of 197 metros, each containing between 50 and 1,000 census tracts.[12] The total number of person-year observations in the estimation sample is 142,692,072.

We sort households into 54 mutually exclusive types: by age of the head of the household (young, middle, old), by housing tenure status (renter, owner), by credit score (low, middle, high), and by race (Black, Hispanic, White/other). Referring to $\varphi^{\tau,\tau'}$, with the exception of race, a household's type can stochastically change over time. Borrowing a method from overlapping generations models in macroeconomics to conserve on state variables, we specify

[10] The data include all individuals with 5 out of the 100 possible terminal 2-digit social security number (SSN) combinations. While the leading SSN digits are based on the birth year/location, the terminal SSN digits are essentially randomly assigned. A SSN is required to be included in the data and we do not capture the experiences of illegal immigrants.

[11] We match census block to census tract using the year-2000 definition of census tracts.

[12] We impose the limitation on the maximum number of census tracts in a metro to keep estimation feasible.

that households age up (i.e. low to middle, middle to high) or die (high to death) with a 5% probability each year. Conditional on age and race, we estimate the annual 6x6 matrix of transition probabilities of housing tenure status and credit score using the CCP data pertaining to our estimation sample.

From the CCP data, we classify a household as young if the age of the household head is between 25-44, middle aged if 45-64, and old if 65 and older. We classify the household as a homeowner if the household has a mortgage and a renter if not. Finally, we classify a household as having a low credit score if the Equifax Risk Score^{TM} of the household head is less than or equal to 599, middle credit score if between 600 and 720 inclusive, and high credit score if greater than or equal to 721.[13]

Before we discuss estimation of the parameters, we first need to explain how we structure the CCP for estimation and document a shortcoming of the data. For each type, we construct an estimate of the probability that ℓ is the end-of-period location given a beginning-of-period location of j. We compute this estimate by pooling all observations across all time periods.

Additionally, while we observe most of the elements of any type τ, we do not directly observe race. We infer information about a household's race from the census block where we first observe the primary sample person in the household.[14] Let the superscript r denote race (r equals w for White/other, b for Black, and h for Hispanic) and define ω_i^r as our estimate of the probability that household i is of race r where $\sum_r \omega_i^r = 1$. For each $r = \{w, b, h\}$, we set ω_i^r for household i equal to that race's share in the census block in which household i is first observed. We then use these probabilities to identify, for each type τ, the conditional probability that a location ℓ' is chosen in metro m given a starting location of j' in metro m that period. Denote $r(\tau)$ as the specific race r associated with type τ. The estimate of that conditional probability is

$$P_m^\tau(\ell' \mid j') = \frac{\sum_t \sum_i \omega_i^{r(\tau)} \mathcal{I}(\ell_{i,t+1} = \ell') \mathcal{I}(j_{i,t} = j')}{\sum_t \sum_i \omega_i^{r(\tau)} \mathcal{I}(j_{i,t} = j')} \tag{24}$$

where $\mathcal{I}(j_{i,t} = j')$ is an indicator that is equal to 1 if household i starts period t in location j' in metro m and is 0 otherwise, and $\mathcal{I}(\ell_{i,t+1} = \ell')$ as an indicator that household i chooses

[13] We keep only households with 4 or fewer adult members. A household is defined as a homeowner based on whether anyone in the household has any type of home loan. The credit score is that of the oldest adult if the household has 2 or fewer adults, and the oldest adult under the age of 65 of there are 3 or 4 adults in the household.

[14] For reference, each census block has about 100 residents and a census tract has about 4,000 residents. If a household is first observed before 2010, then we use racial shares for that household for census blocks from the year-2000 census. If a household is first observed in 2010 or later, we use racial shares for census blocks from the year-2010 census.

period ℓ' in metro m in period t (or, equivalently, starts period $t+1$ in location ℓ').

4.3.2 Potential Implications of Imperfect Measurement of Race

The fact that we do not perfectly observe race suggests that our estimates of choice probabilities by race could be mismeasured. This may ultimately bias our estimates of $\delta^{\tau}_{\ell,m}$. That said, any bias that arises will make location choices look more similar by race than would be estimated if race were perfectly measured.

To see this, consider a simple estimate of the probability location ℓ is chosen by a White household where race is not always measured correctly.[15] This estimate can be written as

$$\hat{P}^{w}(\ell) \;\; = \;\; (1-\phi^{w})\, P^{w}(\ell) + \phi^{w} P^{-w}(\ell)$$

In the equation above, ϕ^w is the fraction of respondents labeled as "White" households that are actually nonwhite, $P^w(\ell)$ is the true probability White households choose tract ℓ and $P^{-w}(\ell)$ is the true probability nonwhite households choose tract ℓ. The estimated probability White households choose tract ℓ, $\hat{P}^w(\ell)$, will be a blended average of the probabilities White and nonwhite households choose tract ℓ. The size of the bias depends on the extent of the mislabeling and the difference of the choice probabilities of White and nonwhite households:

$$\hat{P}^{w}(\ell) \;\; = \;\; P^{w}(\ell) \;\; - \;\; \underbrace{\phi^{w}\left[P^{w}(\ell) - P^{-w}(\ell)\right]}_{\text{bias}}$$

If $P^w(\ell) > P^{-w}(\ell)$, then the bias is negative; estimated choice probabilities by race will appear to be more similar than would be implied if race were perfectly observed.[16]

Some simple math shows that any bias that arises due to mismeasurement is likely to be about one-third the size for White households than for either Black or Hispanic households. The reason is that White households comprise 76 percent of our sample and Black and Hispanic households each account for about 12 percent of our sample. Consider a simple example of a sample of 1000 people with 760 White, 120 Black, and 120 Hispanic. If 10% of Black and 10% of Hispanic households are incorrectly labeled as White, only 24 out of 760 White-labeled households will be mislabeled – about 3 percent. For the overall racial shares in the sample to be accurate, 12 White households will be mislabeled as Black and 12 White

[15] In this simple example we hold all aspects of a household's type other than race fixed.

[16] Obviously, other authors have discussed issues with imputing race in large data sets. One recent proposal for imputing race in administrative data suggests using both full names (or combinations of letters appearing together) and geography: See Cabreros, Agniel, Martino, Damberg, and Elliott (2022). Note that we do not observe names in our data.

households will be mislabeled as Hispanic, 10% each of Black and Hispanic households. Thus, ϕ for White households will be about one-third the size of ϕ for nonwhite households due to simple arithmetic.

As we show later, our current estimates imply many households make choices that suggest they prefer racially segregated neighborhoods. The bias we have discussed pushes estimates away from this finding of homophily, since it shrinks differences across race in estimated location-choice probabilities. In Appendix C, we evaluate the stability of racial shares in neighborhoods in our data by estimating the eigenvalues of our decision model. We show the hypothesis that existing neighborhood racial shares are stable can be rejected because our estimated preferences for homophily are strong. The fact that our estimates may be biased away from finding homophily makes the rejection of stability of neighborhood racial shares even more stark.[17]

4.3.3 Sample Statistics

Table 3 below shows tract level statistics on racial composition, racial mixing, and adjusted rents for the census tracts in our overall estimation sample, the third column, and our estimation sample split into metros with below-median shares of Black households, the fourth and fifth columns, and metros with above-median shares of Black households, the columns six and seven. The first column of each pair (columns four and six) shows results for relatively small metros and the second column shows results for larger metros. We consider two different ways to weight the data. In the first way, marked "tract," we assign to each tract a weight of 1.0. In the second way (marked as "Black pop," "Hisp pop," "White pop"), to compute averages for each tract we assign a weight equal to the population of the indicated race in that tract. For example, focusing on the third row of the table, the value of 0.359 in the third column means that a randomly chosen Black household in our sample on average lives in census tracts with a share of Black households of 35.9%.

We highlight a few results from this table. First, each race tends to have a high proportion of same-race neighbors. On average a Black household lives in census tracts with a share of Black households of 35.9% (as discussed); a Hispanic household lives in census tracts with a Hispanic share of 31.1%; and White households live in census tracts with a White share of 82.4%. Second, the exposure of Black households to White neighbors decreases with the share of Black households in a metro. In metros with a relatively small fraction of Black households, a typical Black household lives in a census tract that is comprised of 68.0% or 78.9% White households (the smaller estimate is from larger metros). These exposure

[17]We discuss our analysis of eigenvalues in more detail at the end of section 5.2.1.

rates fall about 20 percentage points, to 49.7% and 57.2%, in large and small metros with a relatively large fraction of Black households. Finally, composition-adjusted rents are lower in smaller metros compared to larger metros.

Table 3: Statistics on Race and Rent from the Estimation Sample

			Metros with			
			Low Black Share		High Black Share	
Outcome (1)	Weighting (2)	Mean (SD) (3)	Small (4)	Large (5)	Small (6)	Large (7)
Pct. Black	tract	0.152	0.047	0.063	0.241	0.219
		(0.209)	(0.065)	(0.086)	(0.227)	(0.247)
	Black pop.	0.359	0.099	0.138	0.377	0.414
	Hisp pop.	0.105	0.037	0.063	0.189	0.162
	White pop.	0.088	0.036	0.048	0.147	0.119
Pct. Hisp	tract	0.116	0.124	0.163	0.055	0.094
		(0.153)	(0.189)	(0.180)	(0.057)	(0.126)
	Black pop.	0.099	0.112	0.182	0.051	0.089
	Hisp pop.	0.311	0.404	0.346	0.100	0.265
	White pop.	0.089	0.079	0.121	0.050	0.075
Pct. White	tract	0.732	0.829	0.774	0.705	0.687
		(0.244)	(0.198)	(0.202)	(0.231)	(0.268)
	Black pop.	0.543	0.789	0.680	0.572	0.497
	Hisp pop.	0.584	0.560	0.591	0.711	0.573
	White pop.	0.824	0.885	0.831	0.803	0.806
Rent	tract	\$928	\$804	\$1,030	\$762	\$916
		(\$338)	(\$271)	(\$371)	(\$228)	(\$321)
	Black pop.	\$905	\$797	\$1,038	\$752	\$909
	Hisp pop.	\$1,014	\$825	\$1,062	\$868	\$1,026
	White pop.	\$987	\$844	\$1,088	\$821	\$980
Number of tracts		40,556	4,478	12,932	3,256	19,890
Number of metros		197	57	42	42	56

4.3.4 Topography

We use 11 topography variables tabulated to census tracts from Baum-Snow and Han (2022).[18] In that study, the authors construct topographic information using the "Scientific Investigations Map 3085" derived from the US Geological Survey's National Elevation Database.[19] Baum-Snow and Han (2022) write, "This data set uses raster information on

[18] These variables were graciously given to us by Nathaniel "Nate" Baum-Snow and Lu Han.

[19] For reference, see USGS Land Surface Forms available at *https://pubs.usgs.gov/publication/sim3085*

slope and elevation range for all 30X30 meter land pixels within a 0.56 km radius (1 sq. km.) of each pixel to classify it into one of nine categories that describe how flat or hilly the surrounding area is." Each of the 11 variables is the percentage of the census tract characterized by the stated topographical feature, and the sum of the variables in each tract is 1.0. The specific variables are *Flat Plains*, *Smooth Plains*, *Irregular Plains*, *Escarpments*, *Low Hills*, *Hills*, *Breaks-Foothills*, *Low Mountains*, *High Mountains*, *Drainage Channels* and *Everything Else*.

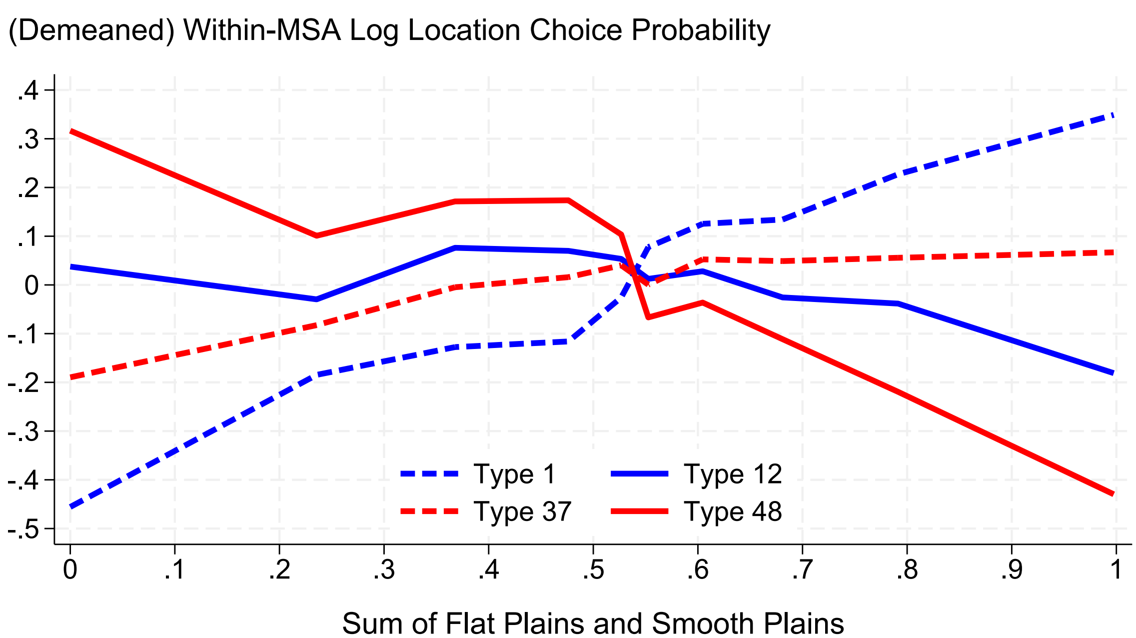
The first two columns of Table 4 report the unconditional means and standard deviations of the 11 topography variables. The third and fourth columns decompose the unconditional standard deviation into the within-metro standard deviation, the third column, and the between-metro standard deviation, the fourth column. The third column shows that there is real variance in topography within metro areas, a condition we exploit in estimation.

Table 4: Means and Standard Deviations of Topography Variables

Variable	Mean	Std. Dev.	Std. Dev. Within	Std. Dev. Between
Flat Plains	0.410	0.423	0.276	0.320
Smooth Plains	0.221	0.262	0.222	0.139
Irregular Plains	0.203	0.265	0.211	0.160
Escarpments	0.006	0.030	0.029	0.010
Low Hills	0.000	0.001	0.001	0.000
Hills	0.032	0.095	0.081	0.050
Breaks-Foothills	0.025	0.088	0.075	0.047
Low Mountains	0.014	0.072	0.067	0.028
High Mountains	0.001	0.015	0.015	0.003
Drainage Channels	0.088	0.096	0.076	0.059
Everything Else	0.001	0.016	0.015	0.004

We now show that the relative, within-metro topography of locations is related to household race and socio-economic status, consistent with results from other papers (Lee and Lin, 2018; Ye and Becker, 2018; Heblich, Trew, and Zylberberg, 2021; Ye and Becker, 2024). Figure 1 below shows how within-metro location choices vary by the the sum of the fraction of the census tract that is made up of *Flat Plains* and *Smooth Plains*. Table 4 shows that these two variables account for 63 percent of a census tract's topography, on average, but that there is significant variation in these variables across-tracts within a metro area. The y-axis shows the log of location choice probabilities after removing metro area fixed effects and the x-axis shows the sum of the *Flat Plains* and *Smooth Plains* variables, also after removing metro area fixed effects. The graph is a binscatter of the data from 10 deciles. The figure shows results for four types of households from our data: young, low-credit score renters, dashed lines (Types 1 and 37), and middle-aged, high-credit score homeowners, solid lines (Types 12 and 48). For both the dashed and solid lines, the blue lines show results for Black households (Types 1 and 12) and the red lines show results for White households (Types 37 and 48). The graph shows how topography and demographics interact: middle-aged, high-credit score homeowners (Types 12 and 48) are more likely to live in hilly, low-*Flat* or *Smooth Plain* areas, than young, low-credit score renters (Types 1 and 37). But, conditional on age and tenure status, White households are more likely to live in hilly areas than Black households: Type 48 is more likely to live in hilly areas than Type 12, and Type 37 is more likely to live in hilly areas than Type 1.

Figure 1: Variation in Location Choice Probabilities by Sum of Flat and Smooth Plains Percentages



Notes: This figure shows a binscatter, across 10 deciles, of the log of location choice probabilities with metro area fixed effects removed (y-axis) against the sum of the *Flat Plains* and *Smooth Plains* Variables, also with metro area fixed effects removed, and recentered to range from 0-1 (x-axis).

In Table 5 below we show coefficient estimates and standard errors for equation (21), the full prediction equation, for the 4 types of households shown in Figure 1. Appendix

B shows estimates and standard errors for all 54 types in our analysis. At the bottom of the table we report the number of observations, overall R^2 including the fixed effects, and within R^2 (R^2 after variation from fixed effects is removed). Three results are apparent from this table. First, relative to the omitted topography variable of *Flat Plains*, most coefficients are statistically significant. Second, within-R^2 values are highest for low-credit score young Black renters, Type 1, and high-credit score middle-aged White homeowners, Type 48. These are the two types of households with likely the largest difference in affluence. Finally, the magnitudes and signs of the coefficients are in alignment with the lines in the graph in Figure 1. Focusing on the column for Type 1, as the percentage of the location (census tract) increases in *Flat Plains*, the omitted variable, the predicted percentage of Type 1 households choosing that location increases.[20] In Figure 1, this outcome is shown by the increasing dashed blue line. In contrast, for Type 48, as the percentage of the location increases in *Flat Plains*, the omitted variable, the predicted percentage of Type 1 households choosing that location *decreases*. This is shown by the declining solid red line in Figure 1.

5 Estimates and Implications

5.1 Estimates

Table 6 provides a summary of our estimates of preferences that household types in our data have over the racial mix of their neighborhood, $\Theta_2^\tau = \{a_1^\tau, \ldots, a_5^\tau\}$, as described in section 4.2. Based on results from Davis, Gregory, Hartley, and Tan (2021), we set a_r^τ equal to 0.243 for all low credit score household types, 0.179 for all middle credit score types, and 0.135 for all high credit score types.[21] This implies that if log rents increase by 0.3 (about a 35% increase in rental prices), that holding all else equal indirect utility declines by 0.073, 0.054 and 0.041 for low-, middle-, and high- credit score types. Keep these values in mind when evaluating the estimated magnitudes of racial preferences reported in Table 6

Column (1) of Table 6 shows the index for household type and (2) reports the percentage of the estimation sample accounted for by that type. Columns (3)-(6) show the race, age (y=young, m=middle-aged, and o=old), homeownership tenure (r=rent, o=own), and bin of credit score (l=low, m=middle, h=high) of the type. Column (7) reports the average share of Black households in the census tracts in which that type tends to live and column (8) shows 0.1 times the average derivative of utility that type would experience from an

[20] The coefficients on all of the variables in the column showing Type 1 results are negative.

[21] The estimates we report in Table 6 almost do not change when we double or halve the values of a_r^τ (not shown).

Table 5: Estimates and Standard Errors for Prediction Equation (21)

Variable	Type 1	Type 12	Type 37	Type 48
Flat Plains (Ref)	-	-	-	-
Smooth Plains	-0.107	0.333	-0.060	0.356
	(0.027)	(0.025)	(0.018)	(0.027)
Irregular Plains	-0.448	0.255	-0.253	0.468
	(0.028)	(0.026)	(0.019)	(0.028)
Escarpments	-0.113	-0.336	-0.237	-0.282
	(0.193)	(0.175)	(0.128)	(0.193)
Low Hills	-9.103	3.480	5.258	17.045
	(8.531)	(7.735)	(5.662)	(8.525)
Hills	-1.572	0.514	-0.160	1.768
	(0.073)	(0.066)	(0.048)	(0.073)
Breaks-Foothills	-1.158	0.453	-0.648	1.038
	(0.078)	(0.071)	(0.052)	(0.078)
Low Mountains	-1.873	-0.170	-0.818	0.941
	(0.091)	(0.082)	(0.060)	(0.091)
High Mountains	-1.080	-0.099	-0.959	0.031
	(0.392)	(0.355)	(0.260)	(0.392)
Drainage Channels	-1.181	0.473	0.065	1.550
	(0.077)	(0.070)	(0.051)	(0.077)
All Else	-4.164	-2.085	-1.666	0.499
	(0.356)	(0.323)	(0.236)	(0.356)
Constant	-1.181	0.473	0.065	1.550
	(0.077)	(0.070)	(0.051)	(0.077)
Obs.	40273	40273	40273	40273
Overall R^2	0.431	0.425	0.579	0.417
Within R^2	0.057	0.011	0.016	0.050

Table 6: Summary of Estimates of Preferences over Race

Type	Sample %	Race	Age	Tenure	Credit	Avg S^b_ℓ	$\frac{1}{10} \cdot \text{Avg}\left(\frac{\Delta\delta_\ell}{\Delta S^b_\ell}\right)$	Avg S^h_ℓ	$\frac{1}{10} \cdot \text{Avg}\left(\frac{\Delta\delta_\ell}{\Delta S^h_\ell}\right)$	$\delta^{95}_\ell - \delta^5_\ell$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
1	2.2		y	r	l	0.42	0.22	0.10	0.15	2.70
2	1.2		y	r	m	0.34	0.26	0.11	0.04	1.98
3	0.6		y	r	h	0.24	0.07	0.11	0.04	0.36
4	0.4		y	o	l	0.36	0.21	0.10	-0.20	2.44
5	0.5		y	o	m	0.27	0.20	0.10	-0.02	1.33
6	0.6		y	o	h	0.17	0.01	0.09	-0.01	0.50
7	1.1		m	r	l	0.45	0.13	0.10	0.03	2.25
8	0.9		m	r	m	0.39	0.17	0.10	-0.03	1.74
9	0.6	Black	m	r	h	0.29	0.05	0.10	0.02	0.36
10	0.4		m	o	l	0.43	0.08	0.09	-0.12	1.48
11	0.6		m	o	m	0.36	0.09	0.10	-0.15	0.93
12	0.9		m	o	h	0.23	0.01	0.09	-0.06	0.28
13	0.3		o	r	l	0.50	0.09	0.09	0.09	1.97
14	0.5		o	r	m	0.46	0.09	0.09	-0.12	1.47
15	1.0		o	r	h	0.36	0.05	0.08	-0.20	0.62
16	0.1		o	o	l	0.52	0.03	0.08	-0.18	1.69
17	0.2		o	o	m	0.46	0.08	0.09	-0.29	1.88
18	0.4		o	o	h	0.31	0.07	0.09	-0.23	0.75
sum	12.3				avg	0.36	0.13	0.10	-0.02	1.53
19	1.6		y	r	l	0.13	0.34	0.35	0.17	1.84
20	1.4		y	r	m	0.11	0.18	0.34	0.15	1.35
21	0.8		y	r	h	0.09	-0.00	0.28	0.08	0.97
22	0.3		y	o	l	0.12	0.30	0.33	0.11	1.65
23	0.5		y	o	m	0.10	0.07	0.30	0.12	0.94
24	0.7		y	o	h	0.08	-0.18	0.23	0.01	1.05
25	0.7		m	r	l	0.14	0.28	0.35	0.11	1.58
26	0.9		m	r	m	0.11	0.20	0.35	0.10	1.25
27	0.8	Hisp	m	r	h	0.09	-0.08	0.30	0.08	0.87
28	0.3		m	o	l	0.13	0.13	0.33	0.03	0.97
29	0.6		m	o	m	0.10	0.03	0.31	0.05	0.57
30	1.1		m	o	h	0.08	-0.13	0.24	-0.00	0.77
31	0.1		o	r	l	0.13	0.27	0.39	0.15	1.51
32	0.3		o	r	m	0.11	0.18	0.38	0.12	1.12
33	1.0		o	r	h	0.09	-0.01	0.30	0.03	0.52
34	0.0		o	o	l	0.14	0.15	0.36	-0.00	0.48
35	0.1		o	o	m	0.11	0.15	0.34	0.10	0.89
36	0.4		o	o	h	0.08	-0.01	0.25	0.07	0.67
sum	11.6				avg	0.10	0.10	0.31	0.09	1.14
37	5.6		y	r	l	0.13	0.21	0.11	0.15	0.77
38	6.1		y	r	m	0.10	0.05	0.11	0.12	0.76
39	5.4		y	r	h	0.08	-0.25	0.09	-0.03	1.63
40	1.5		y	o	l	0.11	0.20	0.10	0.06	0.83
41	3.3		y	o	m	0.09	-0.03	0.09	0.05	1.38
42	6.8		y	o	h	0.07	-0.41	0.08	-0.07	2.03
43	2.7		m	r	l	0.13	0.19	0.11	0.06	0.63
44	4.1		m	r	m	0.10	0.15	0.11	0.05	0.50
45	6.5	White	m	r	h	0.08	-0.27	0.08	-0.04	1.37
46	1.3		m	o	l	0.11	0.09	0.09	-0.04	0.55
47	3.7		m	o	m	0.09	-0.06	0.09	-0.06	0.87
48	11.5		m	o	h	0.07	-0.47	0.08	-0.19	1.73
49	0.5		o	r	l	0.13	0.16	0.11	0.08	0.54
50	1.8		o	r	m	0.10	0.04	0.10	0.07	0.29
51	10.5		o	r	h	0.08	-0.16	0.08	-0.04	1.17
52	0.2		o	o	l	0.12	0.06	0.10	-0.03	0.37
53	0.7		o	o	m	0.10	0.02	0.09	-0.01	0.42
54	4.0		o	o	h	0.07	-0.18	0.08	-0.09	1.25
sum	76.1				avg	0.09	-0.14	0.09	-0.02	1.21

For age: y = young, m = middle-aged, o = old. For tenure: r = renter, o = owner. For credit: l = low, m = middle, h = high.

increase in the share of Black households in the census tracts in which that type tends to live (discussed two paragraphs below). Similarly, column (9) reports the average share of Hispanic households in the census tracts in which which that type tends to live and column (10) shows 0.1 times the average derivative of utility that type would experience from an increase in the share of Hispanic households in the census tracts in which that type tends to live. Note that the values reported in columns (7)-(10) are computed as weighted averages over all tracts in which the type may live, with the weights being the probability that the type lives in the tract.[22]

The top, middle and bottom panels of the table show results for Black, Hispanic, and White types of households, respectively. Focusing on the bottom row of each of the panels, Black households account for 12.3% of our sample, Hispanic households account for 11.6% of our sample, and White households account for 76.1% of our sample. Table 6 shows that same-race sorting is a prominent feature of our data. Columns (7) and (9) show that, on average, Black households live in census tracts that are 36% Black, Hispanic households live in census tracts that are 31% Hispanic and White households live in census tracts that are 82% White.[23]

Columns (8) and (10) show 0.1 times the derivative of utility with respect to exogenous changes in the tract-level Black share (8) or Hispanic share (10). Roughly speaking, given the tracts where each type tends to live, this is the average change in utility resulting from a 10 percentage point increase in the share of Black (column 8) or Hispanic (column 10) households living in each tract. As shown in the bottom row of each panel that summarizes all types of a given race, homophily is a prevalent feature of our data: on average Black households receive additional utility from an increase in Black shares; Hispanic households receive additional utility from an increase in Hispanic shares (as well as Black shares); and White households receive additional utility from an increase in White shares.[24] Note that our results suggest there is considerable within-race heterogeneity in preferences. For example, focusing on White types of households, Type 37 (young, low-credit score renters) accounting for 5.6 percent of the sample experience a relatively large increase in utility as the share of Black households in the neighborhoods where they tend to live increases; whereas Type 48 (middle aged, high-credit score homeowners), accounting for 11.5% of our sample, experience

[22]For example, suppose there are two tracts A and B; and, thinking about column 8, suppose a particular type experiences a -1.0 derivative to utility with respect to the Black share in tract A and a +1.0 derivative to utility with respect to the Black share in tract B. If the probability that that type lives in tract A is 0.20, then we would report a value in column 8 for that type of (1/10) times 0.6, computed as $0.6 = 0.2\,(-1.0) + 0.8\,(1.0)$.

[23]The total shares of Black and White households shown in this table and shown in Table 3 are different because this table uses person weights whereas the other table weights each tract equally to compute total racial shares.

[24]Specifically, utility increases for White households if either the Black or Hispanic shares decrease.

a large decrease in utility as the share of Black households in the neighborhoods where they tend to live increases.

Finally, column (11) illustrates the importance of racial preferences in accounting for location choice in our data. For column 11, we set $a_r^\tau = 0$ and $A_{\ell,m}^\tau = 0$ for all τ and all ℓ and m and then evaluate the the level of utility for each type of household in each census tract; in this calculation, differences in Black and Hispanic shares entirely determine differences in utility across census tracts. For each type, we sort tracts by the level of utility the tract provides; we then report in column (11) the level of utility for the type at the location representing the 95th percentile less the level of utility at the location representing the 5th percentile. When compared to the estimates reported in columns 8 and 10, the average change in utility from a 10 percentage point change in the percentage of households that are Black or Hispanic, these utility differentials attributable entirely to differences in racial composition across neighborhoods are huge.

5.2 Implications: Impact of a Small Policy Shock

Next, we use simulations of the model to study the implications of a somewhat small policy change that simultaneously affects a relatively large number of locations in a metro area. Specifically, for each metro area, we simulate the long-run steady-state predicted response after local governments unexpectedly allow a one-time and immediate 10 percent expansion of all housing developments previously financed using Low Income Housing Tax Credits (LIHTC). We allow this policy to potentially increase racial integration in the new simulated steady state by initially populating the new units with low credit score tenants that have the same demographic mix as the low credit score population of the metro. The thought behind this analysis was to ask if local governments could implement a relatively small place-based policy in many locations at once without causing a lot of disruption. If the policy was sufficiently small, and implemented in enough locations that already had experience with government policy via existing LIHTC developments, perhaps incumbent residents would not move in response to a small influx of new low-credit-score residents that may be of a different average racial mix than existing residents.[25]

Before continuing, we need to define a steady state. A steady state has the features that (i) the mix of household types in each tract is stable (implying shares of Black and Hispanic

[25] Note that Cook, Li, and Binder (2024) show that, in practice and unlike our assumptions in this experiment, Black households are less likely to occupy LIHTC units in "higher opportunity" neighborhoods in a metro area. The point of this experiment is not to exactly predict the racial composition of any new LIHTC units, but rather to study the behavioral response to a policy rule that mechanically tries to force a small amount of extra racial integration into a large number of locations at once.

households in each tract are stable), (ii) the rent in each tract is stable, and (iii) expected future rents and Black and Hispanic shares in each tract are equal to realized rents and shares after all location decisions are made each period.

We implement this counterfactual policy experiment as follows. Denote ΔH as the total number of new LIHTC units that will be built in the metro as a consequence of this policy. In the first step, we remove ΔH housing units (in total) from census tracts that are currently housing low-credit-score households in the metro.[26] Then, in the second step we simulate the model for 5 periods holding $\delta_{\ell,m}$ fixed and $r_{\ell,m}$ fixed for every ℓ in the metro. After these 5 periods, and before adding the new LIHTC units, in each tract and for each type we compute the number of households that need to enter or exit ("births and deaths") such that the data are in a steady state with these ΔH units removed.[27]

Finally, in the third step – the jumping-off point for finding the steady state that occurs after the policy is implemented – we add new LIHTC units in proportion to existing LIHTC units until ΔH units are added. We assume the distribution of household types in these new units is the same as the distribution of household types from the ΔH units removed in the first step. With these three steps, we preserve the metro-wide distribution of household types and maintain the metro-wide aggregate stock of housing, but move ΔH low-credit score households from census tracts without LIHTC units to census tracts with LIHTC units. Importantly, the mix of household types moving into the ΔH new LIHTC units is unlikely to be the same as the mix of household types in the tracts where those units are located.

Once we have taken the three steps listed above, we compute a new steady state for each metro. When households have strong preferences over the demographic composition of their neighborhood, we cannot rule out the possibility that there may be multiple feasible steady states in each metro. We therefore compute a new steady state that is consistent with "myopic" expectations. The steady state we consider – as well as the path to the steady state – implied by this assumption about expectations is unique.

Our algorithm to compute the new steady state with myopic expectations is as follows:

a. For each tract ℓ in metro m, denote the total number of households and the rental price in each tract in the starting steady state as $\mathcal{H}_{\ell,m}$ and $r_{\ell,m}$, respectively.

b. For each tract ℓ in metro m, denote the expected black share and expected hispanic share as $E\left[S^b_{\ell,m}\right]$ and $E\left[S^h_{\ell,m}\right]$. Set both of these equal to their values in the starting steady state.

[26] The housing units are removed in proportion to the low-credit score population of each tract.

[27] Recall that each household in our model faces stochastic transitions over states: age (including death), housing tenure choice, and credit score. For a stable mix of types, at a minimum we need births but also need to account for any other asymmetric type transitions.

c. Given household assumptions of $E\left[S^b_{\ell,m}\right]$, and $E\left[S^h_{\ell,m}\right]$, simulate one period of household decisions, add location- and type-specific births and deaths as computed in the jumping off point of the simulations, and find market clearing rents $r'_{\ell,m}$ and the new housing stock $\mathcal{H}'_{\ell,m}$ in each tract such that the following housing-supply-elasticity holds:

$$\log \mathcal{H}'_{\ell,m} - \log \mathcal{H}_{\ell,m} = \psi_{\ell,m}\left[\log r'_{\ell,m} - \log r_{\ell,m}\right] \tag{25}$$

Where the housing supply in tract ℓ in metro m, $\psi_{\ell,m}$, is given by the estimates in Baum-Snow and Han (2022) with a floor value of 0.025.[28]

d. Given the simulated household location decisions from step [c.], update expected Black and Hispanic shares by setting them equal to realized (simulated) Black and Hispanic shares in each tract, $E\left[S^b_{\ell,m}\right] = S^b_{\ell,m}$ and $E\left[S^h_{\ell,m}\right] = S^h_{\ell,m}$.

e. Repeat steps c and d *until the distribution of types in each tract does not change with one additional iteration.*

To be completely clear, when households solve for their optimal location, they need to know utility today and in the future for all possible locations. The current and future values of $\delta^\tau_{\ell,m}$ in each period have the following as components: (i) fixed amenities $A^\tau_{\ell,m}$, (2) expected racial shares, and (3) actual current market clearing rents given those expected racial shares. At each step in the simulation path, households assume the current and expected value of $\delta^\tau_{\ell,m}$ is fixed at its current value. But, along each step of the simulation path, the value of $\delta^\tau_{\ell,m}$ changes as realized racial shares and market-clearing rents change. Thus, when the model is not in steady state, at each step along the simulation path expected racial shares are not accurate because they are backward looking.

5.2.1 Change to the Racial Composition of Neighborhoods

In the simulations, we track three statistics for each metro. The first statistic we compute is the share of tracts that "tip." We define a tract to have tipped if either the Black share or the Hispanic share changes by 5 percentage points or more in the new steady state relative to the original steady state prior to the policy change. The other two statistics we compute are Black-White and Hispanic-White dissimilarity indices. For each metro m, we compute these indices as

$$\begin{aligned}
\text{Black-White dissimilarity} &= \frac{1}{2}\sum_{\ell\in m}\left|\frac{b_{\ell,m}}{B_m} - \frac{w_{\ell,m}}{W_m}\right| \\
\text{Hispanic-White dissimilarity} &= \frac{1}{2}\sum_{\ell\in m}\left|\frac{h_{\ell,m}}{H_m} - \frac{w_{\ell,m}}{W_m}\right|
\end{aligned}$$

[28] In a handful of tracts, Baum-Snow and Han (2022) estimate a negative supply elasticity.

where $b_{\ell,m}$, $h_{\ell,m}$ and $w_{\ell,m}$ are the numbers of Black, Hispanic, and White households in tract ℓ of metro m and B_m, H_m, and W_m are the numbers of Black, Hispanic, and White households in metro m. If there is perfect mixing of races in each tract, then these indices will equal 0; and if there is perfect segregation then the indices will equal 1.

In Table 7 below, we show the results for 20 representative metros in our sample. We chose these metros using the following process: First, we split our estimation sample of metros into 10 deciles based on number of census tracts. Then, decile by decile we find the median percentage of black households in each metro and further split the metros in each decile into above- and below-median groupings. We then randomly select two metro areas per decile, one in the grouping of below-median-share of Black households one in the above-median grouping. The metro areas in this table and others showing these same metros are sorted in the order they were chosen using this procedure.

Explaining the columns of this table, column (1) shows the (shortened) name of the metro and column (2) shows the percentage of census tracts in that metro with some LIHTC units. Columns (3), (6) and (7), and (10) and (11) show results at our baseline estimate of preferences. Column (3) shows the percentage of tracts that tip after the policy is implemented; columns (6) and (10) show the level of the Black and Hispanic dissimilarity indexes at the jumping-off steady state, respectively; and columns (7) and (11) show the change in those indexes at the new steady state, measured in percentage points. Columns (4), (5), (8), (9), (12), and (13) are discussed two paragraphs below.

The results in this table illustrate that the demographic composition of neighborhoods is not stable, in the sense that a small policy change can cause a huge reshuffling of the population that yields a new steady state that is enormously more segregated than in the current data. When measured at the median metro of the 20 we consider (penultimate row of the table), the policy generates a new steady state where 75% of tracts tip relative to the jumping-off point (column 3), the Black-White dissimilarity index changes by 60.5 percentage points (column 7), and the Hispanic-White dissimilarity index changes by 67.1 percentage points (column 11). These results reinforce the idea that households, on net, want to move to more racially segregated neighborhoods.

Ultimately, neighborhoods are not stable because a sufficient number of households have strong preferences over the racial composition of their neighborhoods. To show this, we rerun the policy experiment after "shrinking" preferences for race. In columns (4), (8) and (12) of the table, we show results after setting $\Theta_2^\tau = \{a_1^\tau, \ldots, a_5^\tau\}$ equal to 0.25 of the baseline estimates, but keep the baseline starting values of $\delta_{\ell,m}^\tau$ unchanged for all types. This keeps baseline preferences for each neighborhood unchanged, but reduces the impact of changes in demographic composition on the desirability of neighborhoods. In columns (5), (9) and

Table 7: Summary of Neighborhood Change in Response to Small Policy Shock

Name	LIHTC %	% Tracts Tip Base	$\frac{1}{4}$	$\frac{1}{8}$	BW Dissim Start	Δ	Δ-$\frac{1}{4}$	Δ-$\frac{1}{8}$	HW Dissim Start	Δ	Δ-$\frac{1}{4}$	Δ-$\frac{1}{8}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Springfield, IL	25	76	18	0	31	60.3	18.5	-0.1	13	72.3	3.4	0.1
Spartanburg, SC	33	100	47	0	20	69.7	28.9	0.1	13	73.0	14.9	0.1
Norwich, CT	16	50	2	0	27	63.3	-0.2	-0.0	23	63.0	-0.1	-0.0
Port St. Lucie, FL	18	92	20	0	27	62.9	10.0	0.0	18	66.8	4.5	0.0
Charleston, WV	29	34	5	0	30	59.2	4.5	-0.1	14	65.7	0.2	0.0
Erie, PA	21	47	3	0	36	53.4	1.5	-0.0	24	56.9	0.2	0.0
Eugene, OR	36	21	0	0	12	66.6	-0.0	-0.0	8	69.9	-0.0	-0.0
Montgomery, AL	40	96	80	11	34	58.3	39.2	3.0	14	74.3	36.4	4.6
Brownsville, TX	22	97	0	0	16	54.1	0.1	-0.1	29	48.2	-0.0	-0.2
Salinas, CA	30	92	5	0	26	61.7	0.9	-0.1	42	41.9	-0.8	-0.2
Utica, NY	21	27	3	0	36	53.2	5.9	-0.1	27	58.8	-2.7	-0.0
Augusta, GA	22	100	86	3	27	65.1	46.3	0.0	14	76.8	28.5	0.2
Lansing, MI	30	74	12	0	33	58.8	2.7	-0.1	21	66.3	1.5	-0.0
Charleston, SC	33	100	90	2	21	70.5	53.9	1.1	13	77.2	32.6	1.3
Knoxville, TN	29	49	9	0	27	64.7	8.0	-0.0	11	73.5	0.5	0.1
Greenville, SC	38	96	69	0	25	65.4	32.8	-0.0	15	73.5	14.7	0.1
Worcester, MA	22	45	0	0	26	60.8	-0.1	-0.0	30	54.0	-0.1	-0.0
Youngstown, OH	26	60	12	0	43	46.9	2.3	-0.0	27	56.1	-1.2	0.0
Albany, NY	20	52	8	0	33	58.4	10.4	-0.1	19	68.9	0.5	0.0
Dayton, OH	38	81	25	0	49	42.4	16.9	-0.2	15	67.5	12.4	0.6
IQR column by column:												
25th Percentile	22	49	3	0	26	58.3	1.5	-0.1	14	58.8	-0.0	-0.0
median	28	75	11	0	27	60.5	7.0	-0.0	16	67.1	0.5	0.0
75th Percentile	33	96	25	0	33	64.7	18.5	0.0	24	73.0	12.4	0.1

(13) of the table, we shrink preferences even more and set $\Theta_2^\tau = \{a_1^\tau, \ldots, a_5^\tau\}$ equal to 0.125 of the baseline estimates. Columns (4) and (5) show that as we shrink preferences for race, the percentage of tracts that tip in the steady state after the policy is implemented falls to zero; and columns (8) and (9) and (12) and (13) show that as we shrink preferences for race, changes in the steady state black-white and hispanic-white dissimilarity indices also fall to zero.

One might wonder if the results we report are due to the nature of the policy experiment we consider. Restated, perhaps it is the case that a different policy experiment evaluated at our baseline preferences would not generate a new steady state that looks so different from the starting point that is based on current data. To investigate this, in every metro area we consider we ask if the model, when evaluated at the starting point of the policy experiments, will return to that same starting point after a tiny shock to the demographic composition of any one neighborhood. We call this evaluation our "eigenvalue analysis." Appendix C details the exact procedure we use to compute eigenvalues (measures of system stability). In summary, in almost every metro we investigate, at least one eigenvalue is larger than 1.0, a condition for system instability. In most metros a substantial fraction of eigenvalues are much larger than 1.0. This implies that in nearly every metro we evaluate, any perturbation to demographic composition in any census tract has the potential to generate a new steady state that looks very different from the current data.

5.2.2 The Speed of Convergence to New Steady State

To understand the importance of expectations in determining the rate at which the model converges to a new steady state, we study two paths for expectations in response to the policy shock. The first assumes household expectations look backwards but are updated every period, i.e. what we have assumed so far to find the new steady state. We call this the "backward-looking path." In the second, we take the steady state arising from the backward-looking path and specify that households assume that particular steady state will occur in every period. We call this the "forward-looking path." This path is also unique, although different from the backward-looking path.

In both the backward- and forward-looking paths, the new steady states are identical but household expectations will be incorrect along the transition path to the new steady state. It turns out that the "miss" between expected and realized racial shares in both cases in each tract will be relatively small. The miss is small because realized racial shares change quite slowly along the backward-looking path and quite rapidly along the forward-looking path.

Table 9 demonstrates the importance of household expectations on the rate of convergence to the new steady state. In this table, we keep track of census tracts in which the Black or Hispanic racial share changed by at least 5 percentage points between steady states at our baseline set of parameter estimates. Columns (3) and (4) report the percentage of tracts in the metro in which the Black (column 3) or Hispanic (column 4) share changed by at least 5 percentage points. Columns (5) and (6) report the median number of years required for 80% of the total change in the tipped racial share to occur for the tracts where the Black-share tips (column 5) and Hispanic share tips (column 6) along the backward-looking path. Columns (7) and (8) report the same for the forward-looking path.

This table shows that when expectations are backward looking, convergence to the new steady state occurs much more slowly than when expectations are forward looking. In the

Table 9: Expectations and Rate of Change Between Steady States

	Total	% Tracts that Tip		Median (w/in CBSA) Years for 80% Convg.			
				Backward Looking		Forward Looking	
CBSA	Tracts	Black	Hispanic	Black	Hispanic	Black	Hispanic
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Springfield, IL	55	75	2	32	68	6	10
Spartanburg, SC	51	100	4	34	80	6	8
Norwich, CT	62	47	27	51	53	5	7
Port St. Lucie, FL	60	90	35	42	41	7	10
Charleston, WV	76	33	1	47	160	6	12
Erie, PA	72	46	8	35	34	5	2
Eugene, OR	78	3	18	158	148	6	14
Montgomery, AL	82	96	1	30	45	5	8
Brownsville, TX	86	1	97	82	46	5	9
Salinas, CA	83	20	89	65	36	8	8
Utica, NY	92	25	10	34	44	5	6
Augusta, GA	95	100	2	33	102	5	6
Lansing, MI	117	73	26	38	51	5	6
Charleston, SC	117	100	5	33	43	5	7
Knoxville, TN	128	48	2	29	50	5	8
Greenville, SC	126	95	6	28	40	6	7
Worcester, MA	163	20	42	51	60	5	8
Youngstown, OH	168	59	7	27	35	5	2
Albany, NY	213	48	10	29	37	5	6
Dayton, OH	208	79	3	26	46	5	8
Median				45.2	61.0	5.5	7.7

backward-looking path, at the median across all the metros we consider, for the median census tract where the Black share tips, 80% of the convergence to the new racial share occurs after 45 years and for the median census track where the Hispanic share tips, 80% of the convergence to the new racial share occurs after 61 years. In contrast, along the forward-looking path these values fall in the range of 5.5 to 7.7 years. In our view, the rate of change along the backward-looking path is so slow that it is conceivable people who are not paying attention are unlikely to notice a change in any given year, even though the change is large and significant when measured over decades.

6 Conclusion

We use a new shift-share IV approach to estimate the extent to which the racial composition of neighborhoods affects household utility and neighborhood choice in a dynamic, forward-looking location-choice model where households care about amenities of neighborhoods as well as the racial composition. We find that many households have very strong preferences for homophily. Same-race preferences are so strong that a relatively small public policy we consider generates a new steady state that involves a radical resorting of the population and even more segregation than currently observed.

References

Aliprantis, D., D. R. Carroll, and E. R. Young (2022): "What explains neighborhood sorting by income and race?," *Journal of Urban Economics*, p. 103508. 4

Almagro, M., E. Chyn, and B. Stuart (2022): "The Effects of Urban Renewal Programs on Gentrification and Inequality," Working Paper. 3, 4

Almagro, M., and T. Dominguez-Iino (2022): "Location Sorting and Endogenous Amenities: Evidence from Amsterdam," University of Chicago Working Paper. 19

Bartik, T. J. (1991): *Who Benefits from State and Local Economic Development Policies?* Upjohn Institute. 1

Baum-Snow, N., and L. Han (2022): "The Microgeography of Housing Supply," Working Paper. 6, 28, 37, 51

Bayer, P., M. D. Casey, W. B. McCartney, J. Orellana-Li, and C. S. Zhang (2022): "Distinguishing Causes of Neighborhood Racial Change: A Nearest Neighbor Design," NBER Working Paper 30487. 3, 4

Bayer, P., F. Ferreira, and R. McMillan (2007): "A Unified Framework for Measuring Preferences for Schools and Neighborhoods," *Journal of Political Economy*, 115(4), 588–638. 1, 3

Bayer, P., R. McMillan, A. Murphy, and C. Timmins (2015): "A Dynamic Model of Demand for Houses and Neighborhoods," Duke University Working Paper. 8, 18

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841–890. 1, 8, 20

Bishop, K. C. (2012): "A Dynamic Model of Location Choice and Hedonic Valuation," Working Paper, Washington University in St. Louis. 1, 20

Board of Governors of the Federal Reserve System (2007): "Report to Congress on Credit Scoring and Its Effects on the Availability and Affordability of Credit," . 24

Borusyak, K., P. Hull, and X. Jaravel (2022): "Quasi-Experimental Shift-Share Research Designs," *The Review of Economic Studies*, 89(1), 181–213. 2, 11

Boustan, L. P. (2010): "Was Postwar Suburbanization White Flight? Evidence from the Black Migration," *The Quarterly Journal of Economics*, 125(1), 417–443. 4

Cabreros, I., D. Agniel, S. C. Martino, C. L. Damberg, and M. N. Elliott (2022): "Predicting Race And Ethnicity To Ensure Equitable Algorithms For Health Care Decision Making," *Health Affairs*, 41(8), 1153–1159. 26

Caetano, G., and V. Maheshri (2021): "A Unified Empirical Framework to Study Segregation," Working Paper. 4, 19, 50

Card, D. (2001): "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration," *Journal of Labor Economics*, 19(1), 22–64. 4

Christensen, P., and C. Timmins (2021): "The damages and distortions from discrimination in the rental housing market," Discussion paper, National Bureau of Economic Research. 4

——— (2022): "Sorting or steering: The effects of housing discrimination on neighborhood choice," *Journal of Political Economy*, 130(8), 2110–2163. 4

Cook, C., P. Z. Li, and A. J. Binder (2024): "Where to Build Affordable Housing? Evaluating the Tradeoffs of Location," Working Paper. 35

Cutler, D. M., E. L. Glaeser, and J. L. Vigdor (1999): "The Rise and Decline of the American Ghetto," *Journal of Political Economy*, 107(3), 455–506. 1

Davis, M. A., J. Gregory, D. A. Hartley, and K. T. K. Tan (2021): "Neighborhood Effects and Housing Vouchers," *Quantitative Economics*, 12(4), 1307–1346. 1, 8, 18, 20, 22, 31

Derenoncourt, E. (2022): "Can You Move to Opportunity? Evidence from the Great Migration," *American Economic Review*, 112(2), 369–408. 4

Ellen, I. G. (2000): *Sharing America's neighborhoods: The prospects for stable racial integration.* Harvard University Press. 4

Ellen, I. G., and G. Torrats-Espinosa (2019): "Gentrification and fair housing: Does gentrification further integration?," *Housing Policy Debate*, 29(5), 835–851. 4

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020): "Bartik Instruments: What, When, Why, and How," *American Economic Review*, 110(8), 2586–2624. 2, 11

Heblich, S., A. Trew, and Y. Zylberberg (2021): "East-Side Story: Historical Pollution and Persistent Neighborhood Sorting," *Journal of Political Economy*, 129(5), 1508–1552. 2, 7, 29

Hotz, V. J., and R. A. Miller (1993): "Conditional Choice Probabilities and the Estimation of Dynamic Models," *The Review of Economic Studies*, 60(3), 497–529. 20

Kennan, J., and J. R. Walker (2011): "The Effect of Expected Income on Individual Migration Decisions," *Econometrica*, 79(1), 211–251. 18

Kuminoff, N. V., V. K. Smith, and C. Timmins (2013): "The New Economics of Equilibrium Sorting and Policy Evaluation Using Housing Markets," *Journal of Economic Literature*, 51(4), 1007–1062. 3

Lee, S., and J. Lin (2018): "Natural Amenities, Neighbourhood Dynamics, and Persistence in the Spatial Distribution of Income," *Review of Economic Studies*, 85(1), 663–694. 29

Neilson, C. A. (2017): "Targeted Vouchers, Competition among Schools, and the Academic Achievement of Poor Students," Working Paper. 20

Schelling, T. C. (1971): "Dynamic Models of Segregation," *The Journal of Mathematical Sociology*, 1(2), 143–186. 3

Shertzer, A., and R. P. Walsh (2019): "Racial Sorting and the Emergence of Segregation in American Cities," *Review of Economics and Statistics*, 101(3), 415–427. 4

Shi, Y., D. Hartley, B. Mazumder, and A. Rajan (2022): "The effects of the Great Migration on urban renewal," *Journal of Public Economics*, 209, 1–12. 4

Stuart, B. A., and E. J. Taylor (2021): "Migration networks and location decisions: Evidence from us mass migration," *American Economic Journal: Applied Economics*, 13(3), 134–175. 13

Ye, V. Y., and C. M. Becker (2018): "The Z-axis: Elevation gradient effects in Urban America," *Regional Science and Urban Economics*, 70, 312–329. 2, 7, 29

——— (2024): "Moving mountains: Geography, neighborhood sorting, and spatial income segregation," *Journal of Regional Science*, In Press. 2, 7, 29

Not For Publication Internet Appendix

A Hotz-Miller Expression for Continuation Values

In our estimation sample time window, we assume that $\delta^{\tau}_{\ell,m,t}$ is fixed for each ℓ, m and τ, such that time subscripts can be removed. Also in what follows, we will hold the metro fixed such that the metro subscript m can be removed. When the ϵ are assumed to be drawn i.i.d. from the Type 1 Extreme Value Distribution, the expected value function $E\left[V^{\tau}(j)\right]$ has the functional form

$$E\left[V^{\tau}(j)\right] \;=\; \log\left\{\sum_{\ell=1}^{J}\exp\widetilde{V}^{\tau}(\ell\mid j)\right\}+\zeta \tag{26}$$

where ζ is equal to Euler's constant and

$$\widetilde{V}^{\tau}(\ell\mid j) \;=\; \delta^{\tau}_{\ell} \;-\; \kappa^{\tau}\cdot 1_{\ell\neq j} \;+\; \beta\sum_{\tau'}\varphi^{\tau,\tau'}E\left[V^{\tau'}(\ell)\right] \tag{27}$$

That is, the tilde symbol signifies that the shock ϵ_{ℓ} has been omitted.

Now, we show that the log probabilities that choices are observed are simple functions of model parameters δ^{τ}_{ℓ}, κ^{τ}, β and of observed choice probabilities. To see this, start by noting the log of the probability that location ℓ is chosen by type τ given a current location of j, call it $p^{\tau}(\ell\mid j)$, has the solution

$$p^{\tau}(\ell\mid j) \;=\; \widetilde{V}^{\tau}(\ell\mid j) \;-\; \log\left\{\sum_{\ell'=1}^{J}\exp\left[\widetilde{V}^{\tau}(\ell'\mid j)\right]\right\} \tag{28}$$

Denote ℓ_0 as a reference tract. Subtract and add $\widetilde{V}^{\tau}(\ell_0\mid j)$ to the right-hand side of the above to derive

$$p^{\tau}(\ell\mid j) \;=\; \widetilde{V}^{\tau}(\ell\mid j)-\widetilde{V}^{\tau}(\ell_0\mid j) \;-\; \log\left\{\sum_{\ell'=1}^{J}\exp\left[\widetilde{V}^{\tau}(\ell'\mid j)-\widetilde{V}^{\tau}(\ell_0\mid j)\right]\right\} \tag{29}$$

Note that equation (27) implies

$$\begin{aligned}&\widetilde{V}^{\tau}(\ell\mid j)-\widetilde{V}^{\tau}(\ell_0\mid j)\\ &=\delta^{\tau}_{\ell}-\delta^{\tau}_{\ell_0} \;-\; \kappa^{\tau}\left[1_{\ell\neq j}-1_{\ell_0\neq j}\right] \;+\; \beta\sum_{\tau'}\varphi^{\tau,\tau'}\left\{E\left[V^{\tau'}(\ell)\right]-E\left[V^{\tau'}(\ell_0)\right]\right\}\end{aligned} \tag{30}$$

But from equation (26),

$$E\left[V^{\tau'}(\ell)\right]-E\left[V^{\tau'}(\ell_0)\right] \;=\; \log\left\{\sum_{\ell'=1}^{J}\exp\widetilde{V}^{\tau'}(\ell'\mid\ell)\right\}-\log\left\{\sum_{\ell'=1}^{J}\exp\widetilde{V}^{\tau'}(\ell'\mid\ell_0)\right\}$$

Now note that equation (28) implies

$$\begin{aligned} p^{\tau'}(\ell_0\mid\ell) &= \widetilde{V}^{\tau'}(\ell_0\mid\ell) \;-\; \log\left\{\sum_{\ell'=1}^{J}\exp\left[\widetilde{V}^{\tau'}(\ell'\mid\ell)\right]\right\} \\ p^{\tau'}(\ell_0\mid\ell_0) &= \widetilde{V}^{\tau'}(\ell_0\mid\ell_0) \;-\; \log\left\{\sum_{\ell'=1}^{J}\exp\left[\widetilde{V}^{\tau'}(\ell'\mid\ell_0)\right]\right\} \end{aligned}$$

and thus

$$\log\left\{\sum_{\ell'=1}^{J}\exp\left[\widetilde{V}^{\tau'}(\ell'\mid\ell)\right]\right\}-\log\left\{\sum_{\ell'=1}^{J}\exp\left[\widetilde{V}^{\tau'}(\ell'\mid\ell_0)\right]\right\}$$

is equal to

$$\begin{aligned} & \widetilde{V}^{\tau'}(\ell_0\mid\ell)-\widetilde{V}^{\tau'}(\ell_0\mid\ell_0) \;-\; \left[p^{\tau'}(\ell_0\mid\ell)-p^{\tau'}(\ell_0\mid\ell_0)\right] \\ = & \quad -\kappa^{\tau'}\cdot 1_{\ell\neq\ell_0} \;-\; \left[p^{\tau'}(\ell_0\mid\ell)-p^{\tau'}(\ell_0\mid\ell_0)\right] \end{aligned}$$

The last line is quickly derived from equation (27). Therefore,

$$E\left[V^{\tau'}(\ell)\right]-E\left[V^{\tau'}(\ell_0)\right] \;=\; -\left[p^{\tau'}(\ell_0\mid\ell)-p^{\tau'}(\ell_0\mid\ell_0)+\kappa^{\tau'}\cdot 1_{\ell\neq\ell_0}\right]$$

and equation (30) has the expression

$$\begin{aligned} & \widetilde{V}^{\tau}(\ell\mid j)-\widetilde{V}^{\tau}(\ell_0\mid j) \\ & =\delta_{\ell}^{\tau}-\delta_{\ell_0}^{\tau} \;-\; \kappa^{\tau}\left[1_{\ell\neq j}-1_{\ell_0\neq j}\right] \;-\; \beta\sum_{\tau'}\varphi^{\tau,\tau'}\left[p^{\tau'}(\ell_0\mid\ell)-p^{\tau'}(\ell_0\mid\ell_0)+\kappa^{\tau'}\cdot 1_{\ell\neq\ell_0}\right] \end{aligned} \tag{31}$$

Due to data limitations we discuss in the paper, we combine data across multiple years when estimating probabilities and preference parameters. For this reason, we assume value functions and expectations are fixed in our sample period.

B Prediction Equation Estimates and Standard Errors

Coefficient Estimates and Standard Errors
Black Households

Type	Const	Smooth Plains	Irreg. Plains	Escarf-Ments	Low Hills	Hills	Breaks-Foothills	Low Mntns.	High Mntns.	Drainage Channels	All Else
1	-1.181	-0.107	-0.448	-0.113	-9.103	-1.572	-1.158	-1.873	-1.080	-1.181	-4.164
yng,rnt,low	(0.077)	(0.027)	(0.028)	(0.193)	(8.531)	(0.073)	(0.078)	(0.091)	(0.392)	(0.077)	(0.356)
2	-0.919	0.003	-0.250	-0.064	-1.675	-1.072	-0.851	-1.501	-1.015	-0.919	-3.685
yng,rnt,med	(0.066)	(0.023)	(0.024)	(0.165)	(7.283)	(0.062)	(0.067)	(0.078)	(0.334)	(0.066)	(0.304)
3	-0.404	0.194	0.077	0.007	0.241	-0.372	-0.330	-1.013	-0.919	-0.404	-2.877
yng,rnt,high	(0.061)	(0.022)	(0.023)	(0.154)	(6.800)	(0.058)	(0.062)	(0.072)	(0.312)	(0.061)	(0.284)
4	-0.477	-0.062	-0.388	-0.746	-0.796	-0.854	-0.680	-1.276	-1.044	-0.477	-4.662
yng,own,low	(0.082)	(0.029)	(0.030)	(0.207)	(9.148)	(0.078)	(0.084)	(0.098)	(0.420)	(0.082)	(0.382)
5	-0.066	0.088	-0.181	-0.750	6.138	-0.266	-0.280	-0.992	-0.950	-0.066	-3.987
yng,own,med	(0.072)	(0.025)	(0.026)	(0.181)	(7.962)	(0.068)	(0.073)	(0.085)	(0.366)	(0.072)	(0.332)
6	0.579	0.317	0.226	-0.420	7.439	0.623	0.339	-0.438	-0.727	0.579	-3.130
yng,own,high	(0.073)	(0.026)	(0.027)	(0.183)	(8.085)	(0.069)	(0.074)	(0.086)	(0.371)	(0.073)	(0.337)
7	-1.178	-0.103	-0.448	-0.139	-8.747	-1.555	-1.137	-1.622	-0.924	-1.178	-3.757
mid,rnt,low	(0.080)	(0.028)	(0.030)	(0.201)	(8.868)	(0.075)	(0.081)	(0.095)	(0.407)	(0.080)	(0.370)
8	-0.859	-0.005	-0.290	-0.242	-6.149	-1.053	-0.843	-1.235	-0.533	-0.859	-3.237
mid,rnt,med	(0.069)	(0.024)	(0.026)	(0.174)	(7.669)	(0.065)	(0.070)	(0.082)	(0.352)	(0.069)	(0.320)
9	-0.379	0.182	0.069	0.037	1.248	-0.258	-0.222	-0.641	-0.123	-0.379	-2.439
mid,rnt,high	(0.062)	(0.022)	(0.023)	(0.157)	(6.921)	(0.059)	(0.064)	(0.074)	(0.318)	(0.062)	(0.289)
10	-0.546	0.010	-0.328	-0.573	-12.456	-0.860	-0.470	-1.032	-0.752	-0.546	-4.829
mid,own,low	(0.087)	(0.031)	(0.032)	(0.218)	(9.617)	(0.082)	(0.088)	(0.103)	(0.442)	(0.087)	(0.401)
11	-0.271	0.107	-0.164	-0.608	-4.519	-0.387	-0.208	-0.777	-0.630	-0.271	-4.092
mid,own,med	(0.075)	(0.026)	(0.028)	(0.188)	(8.278)	(0.070)	(0.076)	(0.088)	(0.380)	(0.075)	(0.345)
12	0.473	0.333	0.255	-0.336	3.480	0.514	0.453	-0.170	-0.099	0.473	-2.085
mid,own,high	(0.070)	(0.025)	(0.026)	(0.175)	(7.735)	(0.066)	(0.071)	(0.082)	(0.355)	(0.070)	(0.323)
13	-1.363	-0.093	-0.450	-0.059	-24.128	-1.673	-1.117	-1.678	-0.620	-1.363	-3.767
old,rnt,low	(0.096)	(0.034)	(0.036)	(0.242)	(10.668)	(0.091)	(0.098)	(0.114)	(0.490)	(0.096)	(0.445)
14	-1.027	0.005	-0.279	-0.023	-12.208	-1.181	-0.758	-1.202	-0.480	-1.027	-3.454
old,rnt,med	(0.082)	(0.029)	(0.030)	(0.206)	(9.074)	(0.077)	(0.083)	(0.097)	(0.417)	(0.082)	(0.379)
15	-0.597	0.194	0.062	-0.047	-3.021	-0.500	-0.264	-0.746	-0.338	-0.597	-2.241
old,rnt,high	(0.070)	(0.025)	(0.026)	(0.177)	(7.818)	(0.066)	(0.072)	(0.083)	(0.359)	(0.070)	(0.326)
16	-1.034	0.010	-0.297	-0.757	-22.886	-1.039	-0.352	-1.013	-0.975	-1.034	-4.252
old,own,low	(0.109)	(0.039)	(0.040)	(0.274)	(12.065)	(0.103)	(0.111)	(0.129)	(0.554)	(0.109)	(0.503)
17	-0.727	0.098	-0.198	-0.482	-13.229	-0.734	-0.268	-0.841	-0.480	-0.727	-3.577
old,own,med	(0.094)	(0.033)	(0.035)	(0.237)	(10.445)	(0.089)	(0.096)	(0.111)	(0.480)	(0.094)	(0.436)
18	0.106	0.319	0.216	-0.118	-1.416	0.151	0.339	-0.220	-0.240	0.106	-1.638
old,own,high	(0.080)	(0.028)	(0.029)	(0.201)	(8.854)	(0.075)	(0.081)	(0.094)	(0.407)	(0.080)	(0.369)

Notes: This table shows the coefficient estimates and standard errors for the prediction equation (21) of location choice on topographic variables by household type. Legend: yng = young, mid = middle aged, old = old aged, rnt = renter, own = owner, low = low credit score, med = medium credit score, high = high credit score.

Coefficient Estimates and Standard Errors
Hispanic Households

Type	Const	Smooth Plains	Irreg. Plains	Escarf- Ments	Low Hills	Hills	Breaks-Foothills	Low Mntns.	High Mntns.	Drainage Channels	All Else
19	-0.710	-0.198	-0.529	-0.592	-2.122	-1.262	-1.464	-1.699	-1.481	-0.710	-3.811
yng,rnt,low	(0.067)	(0.024)	(0.025)	(0.168)	(7.389)	(0.063)	(0.068)	(0.079)	(0.339)	(0.067)	(0.308)
20	-0.415	-0.113	-0.353	-0.593	4.062	-0.775	-1.203	-1.396	-1.548	-0.415	-3.363
yng,rnt,med	(0.061)	(0.022)	(0.023)	(0.154)	(6.802)	(0.058)	(0.062)	(0.073)	(0.312)	(0.061)	(0.284)
21	0.070	0.064	-0.034	-0.505	11.286	-0.029	-0.649	-0.970	-1.443	0.070	-2.426
yng,rnt,high	(0.063)	(0.022)	(0.023)	(0.159)	(7.032)	(0.060)	(0.065)	(0.075)	(0.323)	(0.063)	(0.293)
22	-0.076	-0.188	-0.516	-1.310	3.385	-0.556	-0.934	-1.189	-1.784	-0.076	-4.512
yng,own,low	(0.078)	(0.028)	(0.029)	(0.197)	(8.696)	(0.074)	(0.080)	(0.093)	(0.399)	(0.078)	(0.363)
23	0.308	-0.022	-0.297	-1.120	12.192	-0.011	-0.578	-0.933	-1.664	0.308	-3.998
yng,own,med	(0.072)	(0.026)	(0.027)	(0.182)	(8.041)	(0.068)	(0.074)	(0.086)	(0.369)	(0.072)	(0.336)
24	0.762	0.177	0.114	-0.660	13.861	0.751	0.004	-0.471	-1.216	0.762	-2.821
yng,own,high	(0.076)	(0.027)	(0.028)	(0.191)	(8.400)	(0.071)	(0.077)	(0.090)	(0.386)	(0.076)	(0.350)
25	-0.664	-0.202	-0.539	-0.716	0.345	-1.159	-1.397	-1.349	-1.134	-0.664	-3.089
mid,rnt,low	(0.068)	(0.024)	(0.025)	(0.172)	(7.606)	(0.065)	(0.070)	(0.081)	(0.349)	(0.068)	(0.317)
26	-0.248	-0.134	-0.383	-0.836	3.616	-0.648	-1.125	-1.034	-1.108	-0.248	-2.698
mid,rnt,med	(0.061)	(0.022)	(0.023)	(0.154)	(6.789)	(0.058)	(0.062)	(0.072)	(0.312)	(0.061)	(0.283)
27	0.232	0.053	-0.052	-0.652	13.993	0.213	-0.505	-0.464	-0.974	0.232	-1.599
mid,rnt,high	(0.062)	(0.022)	(0.023)	(0.155)	(6.834)	(0.058)	(0.063)	(0.073)	(0.314)	(0.062)	(0.285)
28	-0.041	-0.093	-0.431	-1.289	-5.961	-0.468	-0.821	-0.836	-1.262	-0.041	-4.065
mid,own,low	(0.077)	(0.027)	(0.029)	(0.195)	(8.584)	(0.073)	(0.079)	(0.091)	(0.394)	(0.077)	(0.358)
29	0.250	-0.019	-0.243	-1.085	8.112	0.014	-0.527	-0.583	-1.149	0.250	-3.160
mid,own,med	(0.070)	(0.025)	(0.026)	(0.175)	(7.725)	(0.066)	(0.071)	(0.082)	(0.355)	(0.070)	(0.322)
30	0.827	0.204	0.145	-0.708	13.207	0.845	0.101	-0.056	-0.871	0.827	-1.427
mid,own,high	(0.071)	(0.025)	(0.026)	(0.178)	(7.870)	(0.067)	(0.072)	(0.084)	(0.361)	(0.071)	(0.328)
31	-0.838	-0.229	-0.570	-0.858	-4.361	-1.227	-1.493	-1.579	-1.209	-0.838	-3.197
old,rnt,low	(0.086)	(0.031)	(0.032)	(0.217)	(9.574)	(0.081)	(0.088)	(0.102)	(0.440)	(0.086)	(0.399)
32	-0.334	-0.135	-0.398	-0.732	2.903	-0.730	-1.058	-1.118	-0.720	-0.334	-2.377
old,rnt,med	(0.069)	(0.025)	(0.026)	(0.175)	(7.699)	(0.065)	(0.071)	(0.082)	(0.354)	(0.069)	(0.321)
33	0.193	0.046	-0.050	-0.661	10.207	0.035	-0.543	-0.542	-0.939	0.193	-1.300
old,rnt,high	(0.065)	(0.023)	(0.024)	(0.164)	(7.226)	(0.061)	(0.066)	(0.077)	(0.332)	(0.065)	(0.301)
34	-0.360	-0.096	-0.383	-0.975	16.188	-0.581	-0.833	-0.858	-1.525	-0.360	-3.695
old,own,low	(0.104)	(0.037)	(0.038)	(0.261)	(11.519)	(0.098)	(0.106)	(0.123)	(0.529)	(0.104)	(0.481)
35	-0.060	-0.015	-0.297	-1.012	4.905	-0.275	-0.634	-0.650	-0.812	-0.060	-2.938
old,own,med	(0.085)	(0.030)	(0.031)	(0.213)	(9.402)	(0.080)	(0.086)	(0.100)	(0.432)	(0.085)	(0.392)
36	0.679	0.196	0.126	-0.723	16.122	0.636	0.052	-0.018	-0.636	0.679	-0.833
old,own,high	(0.076)	(0.027)	(0.028)	(0.190)	(8.386)	(0.071)	(0.077)	(0.089)	(0.385)	(0.076)	(0.350)

Notes: This table shows the coefficient estimates and standard errors for the prediction equation (21) of location choice on topographic variables by household type. Legend: yng = young, mid = middle aged, old = old aged, rnt = renter, own = owner, low = low credit score, med = medium credit score, high = high credit score.

Coefficient Estimates and Standard Errors
White Households

Type	Const	Smooth Plains	Irreg. Plains	Escarf-Ments	Low Hills	Hills	Breaks-Foothills	Low Mntns.	High Mntns.	Drainage Channels	All Else
37	0.065	-0.060	-0.253	-0.237	5.258	-0.160	-0.648	-0.818	-0.959	0.065	-1.666
yng,rnt,low	(0.051)	(0.018)	(0.019)	(0.128)	(5.662)	(0.048)	(0.052)	(0.060)	(0.260)	(0.051)	(0.236)
38	0.415	0.073	-0.008	-0.074	8.580	0.393	-0.256	-0.386	-0.813	0.415	-1.033
yng,rnt,med	(0.053)	(0.019)	(0.020)	(0.134)	(5.900)	(0.050)	(0.054)	(0.063)	(0.271)	(0.053)	(0.246)
39	0.793	0.250	0.308	0.110	12.604	1.032	0.311	0.069	-0.592	0.793	-0.272
yng,rnt,high	(0.065)	(0.023)	(0.024)	(0.163)	(7.170)	(0.061)	(0.066)	(0.076)	(0.329)	(0.065)	(0.299)
40	0.732	-0.039	-0.216	-0.660	9.012	0.446	-0.149	-0.326	-1.169	0.732	-2.407
yng,own,low	(0.068)	(0.024)	(0.025)	(0.172)	(7.581)	(0.064)	(0.070)	(0.081)	(0.348)	(0.068)	(0.316)
41	1.033	0.129	0.040	-0.671	12.070	0.940	0.283	-0.009	-1.006	1.033	-1.620
yng,own,med	(0.069)	(0.025)	(0.026)	(0.174)	(7.679)	(0.065)	(0.070)	(0.082)	(0.353)	(0.069)	(0.320)
42	1.341	0.325	0.391	-0.310	16.693	1.550	0.832	0.420	-0.636	1.341	-0.769
yng,own,high	(0.078)	(0.028)	(0.029)	(0.197)	(8.690)	(0.074)	(0.080)	(0.093)	(0.399)	(0.078)	(0.363)
43	0.179	-0.063	-0.241	-0.300	3.642	-0.004	-0.543	-0.473	-0.499	0.179	-1.134
mid,rnt,low	(0.051)	(0.018)	(0.019)	(0.129)	(5.686)	(0.048)	(0.052)	(0.061)	(0.261)	(0.051)	(0.237)
44	0.634	0.037	-0.042	-0.275	10.530	0.544	-0.155	-0.012	-0.351	0.634	-0.446
mid,rnt,med	(0.052)	(0.018)	(0.019)	(0.131)	(5.776)	(0.049)	(0.053)	(0.062)	(0.265)	(0.052)	(0.241)
45	1.085	0.221	0.309	-0.046	18.836	1.351	0.554	0.638	0.102	1.085	0.724
mid,rnt,high	(0.066)	(0.023)	(0.024)	(0.166)	(7.299)	(0.062)	(0.067)	(0.078)	(0.335)	(0.066)	(0.305)
46	0.909	0.052	-0.088	-0.662	2.296	0.682	0.137	0.176	-0.608	0.909	-1.756
mid,own,low	(0.066)	(0.023)	(0.024)	(0.165)	(7.291)	(0.062)	(0.067)	(0.078)	(0.335)	(0.066)	(0.304)
47	1.167	0.161	0.123	-0.548	10.307	1.135	0.450	0.450	-0.390	1.167	-0.818
mid,own,med	(0.066)	(0.024)	(0.025)	(0.167)	(7.365)	(0.063)	(0.068)	(0.078)	(0.338)	(0.066)	(0.307)
48	1.550	0.356	0.468	-0.282	17.045	1.768	1.038	0.941	0.031	1.550	0.499
mid,own,high	(0.077)	(0.027)	(0.028)	(0.193)	(8.525)	(0.073)	(0.078)	(0.091)	(0.392)	(0.077)	(0.356)
49	0.191	-0.076	-0.274	-0.536	6.870	-0.121	-0.672	-0.565	-0.447	0.191	-1.078
old,rnt,low	(0.066)	(0.023)	(0.024)	(0.166)	(7.299)	(0.062)	(0.067)	(0.078)	(0.335)	(0.066)	(0.305)
50	0.642	0.057	-0.010	-0.276	10.468	0.502	-0.132	0.048	-0.175	0.642	0.027
old,rnt,med	(0.058)	(0.020)	(0.021)	(0.145)	(6.391)	(0.054)	(0.059)	(0.068)	(0.294)	(0.058)	(0.267)
51	1.085	0.238	0.336	-0.104	18.406	1.167	0.508	0.536	-0.005	1.085	1.061
old,rnt,high	(0.071)	(0.025)	(0.026)	(0.178)	(7.856)	(0.067)	(0.072)	(0.084)	(0.361)	(0.071)	(0.328)
52	0.611	0.014	-0.113	-0.740	11.761	0.483	0.074	0.108	-1.511	0.611	-1.359
old,own,low	(0.089)	(0.032)	(0.033)	(0.225)	(9.929)	(0.084)	(0.091)	(0.106)	(0.456)	(0.089)	(0.414)
53	0.919	0.141	0.090	-0.390	8.018	0.931	0.377	0.467	-0.303	0.919	-0.132
old,own,med	(0.071)	(0.025)	(0.026)	(0.179)	(7.902)	(0.067)	(0.073)	(0.084)	(0.363)	(0.071)	(0.330)
54	1.439	0.359	0.480	-0.217	18.941	1.605	0.948	0.995	0.301	1.439	1.570
old,own,high	(0.077)	(0.027)	(0.029)	(0.195)	(8.601)	(0.073)	(0.079)	(0.092)	(0.395)	(0.077)	(0.359)

Notes: This table shows the coefficient estimates and standard errors for the prediction equation (21) of location choice on topographic variables by household type. Legend: yng = young, mid = middle aged, old = old aged, rnt = renter, own = owner, low = low credit score, med = medium credit score, high = high credit score.

C Eigenvalue Analysis

We wish to understand if our estimates imply that the current demographic composition of neighborhoods is stable. This has been studied before by Caetano and Maheshri (2021) and others, but our methods and definition of stability are going to be different. We begin by introducing some notation and defining what we mean by stability. For a given metro m with J_m total tracts, denote $\mathcal{T}$ as a $2J_m \times 1$ vector comprised of starting values of expectations of racial shares, $E\left[S^b_{\ell,m}\right]$ and $E\left[S^h_{\ell,m}\right]$ for all tracts. Let $g(\mathcal{T})$ be an expectations-generating function produced by our model that takes as a starting input $\mathcal{T}$ and produces a different vector of expectations $\mathcal{T}'$,

$$\mathcal{T}' = g(\mathcal{T}).$$

We define a steady state of g as a vector of expectations $\mathcal{T}^*$ that generates, via g, an identical set of expectations, i.e.

$$\mathcal{T}^* = g(\mathcal{T}^*).$$

Before describing how we compute $g(\mathcal{T})$, we now define a steady state that is consistent with the data in our estimation sample for each metro. We start with the distribution of types by tract implied by our estimation sample and then simulate the model for 5 periods, our "burn in" period. During these 5 periods, we assume each household's type stays fixed. During the burn in period, we hold $\delta^{\tau}_{\ell,m}$ fixed for all types and all tracts in all metros. We use a 5-period burn in to ensure all types populate all tracts in our baseline steady state implied by the data.[29] After the burn-in, we use the resulting distribution of types by tract to compute our baseline vector for $\mathcal{T}$, $E\left[S^b_{\ell,m}\right] = S^b_{\ell,m}$ and $E\left[S^h_{\ell,m}\right] = S^h_{\ell,m}$ for all ℓ and m.

Next, we compute the distribution of types across all tracts that results after running the decision model for one period such that all location choices are made and all types probabilistically evolve. For each tract, we compute the required additions ("births") or subtractions ("deaths") of the population of each type such that the resulting measures of household types in each tract after all decisions are made and all types have stochastically evolved is constant in all tracts. The addition of type-specific births and deaths to each tract guarantees that the model-predicted distribution of types across tracts is stable and the vector $\mathcal{T}^*$ reflecting our data is a steady state. That is, the decisions implied by the model are consistent with expectations households have over racial shares and rental prices in each tract.

[29] The burn-in period smoothes through sampling variability in the data.

We now describe the $g(\mathcal{T})$ function that we use to predict how expectations evolve given any starting set of expectations $\mathcal{T}$. To start, denote the total number of households and the rental price in each tract in the data as $\mathcal{H}_{\ell,m}$ and $r_{\ell,m}$, respectively. Then, we compute $g(\mathcal{T})$ as follows:

1. Denote the guess of new rental prices $r'_{\ell,m}$.
2. Using equation (15), adjust $\delta^{\tau}_{\ell,m}$ appropriately for all ℓ, m, and τ given the values of $E\left[S^b_{\ell,m}\right]$ and $E\left[S^h_{\ell,m}\right]$ from $\mathcal{T}$ and the guess $r'_{\ell,m}$, holding exogenous amenities $A^{\tau}_{\ell,m}$ fixed. Households assume this new value of $\delta^{\tau}_{\ell,m}$ is fixed forever when making decisions.
3. Simulate the model 99 periods and compute new housing demand in each tract in each metro, $\mathcal{H}'_{\ell,m}$.
4. Update the guess of rental prices and repeat steps 2-3 until rental prices in each tract clear markets to satisfy

$$\log \mathcal{H}'_{\ell,m} - \log \mathcal{H}_{\ell,m} \quad = \quad \psi_{\ell,m}\left[\log r'_{\ell,m} - \log r_{\ell,m}\right]$$

   The housing supply elasticity in each tract ℓ in each metro m, $\psi_{\ell,m}$, is given by the estimates in Baum-Snow and Han (2022) with a floor value of 0.025.[30]
5. Once we know rental prices $r'_{\ell,m}$ that clear housing markets given values of $E\left[S^b_{\ell,m}\right]$ and $E\left[S^h_{\ell,m}\right]$ from $\mathcal{T}$, compute simulated Black and Hispanic shares in each tract and call these $S^{b'}_{\ell,m}$ and $S^{h'}_{\ell,m}$.
6. Set the elements of $\mathcal{T}'$ equal to $S^{b'}_{\ell,m}$ and $S^{h'}_{\ell,m}$.

Given our procedure to compute $g(\mathcal{T})$, we test the stability of the steady state implied by the data by computing the eigenvalues and eigenvectors of the model at the steady state. To see why this is useful, suppose we perturb expectations of racial shares at the steady state – call these perturbed expectations as $\mathcal{T}'$ – and then measure how expectations evolve from this perturbed starting point, i.e. $T'' = g(\mathcal{T}')$. We can do this with a first-order linear approximation:

$$g(\mathcal{T}') - g(T^*) \quad \approx \quad \mathcal{G}\cdot\left[\mathcal{T}' - \mathcal{T}^*\right]$$

where $\mathcal{G}$ is a $2J_m$ by $2J_m$ vector of derivatives of g evaluated at $\mathcal{T}^*$. Once we make appropriate substitutions, we get

$$\left[\mathcal{T}'' - \mathcal{T}^*\right] \quad \approx \quad \mathcal{G}\cdot\left[\mathcal{T}' - \mathcal{T}^*\right]$$

We compute the elements of $\mathcal{G}$ at $\mathcal{T}^*$ using numerical derivatives. Specifically, define $\widetilde{\mathcal{T}}^*_i$

[30] In a handful of tracts, Baum-Snow and Han (2022) estimate a negative supply elasticity.

as equal to $\mathcal{T}^*$ in all elements except for the i^{th} element which we perturb by Δ_i units.[31] We set the i^{th} column of $\mathcal{G}$ equal to $\left[g\left(\widetilde{\mathcal{T}}_i^*\right) - \mathcal{T}^*\right]/\Delta_i$. For each metro, we repeat this computation for all $i = 1, \ldots, 2J_m$ elements of $\mathcal{T}^*$ to populate all the columns of $\mathcal{G}$.

Once we have an estimate of $\mathcal{G}$, we compute its eigenvalues to determine whether the expectations of racial shares move away from or return to the steady-state expectations implied by the data in response to a tiny perturbation to expectations. In other words, we ask if the system predicts expectations return to $\mathcal{T}^*$ if we start our model using expectations that are nearly but not exactly identical to $\mathcal{T}^*$. If all the eigenvalues of $\mathcal{G}$ are less than 1, the expectations converge back to the steady state; if at least one eigenvalue is greater than 1, expectations do not converge back to the starting point and if this is the case, we say the steady state implied by the data is not stable.

The results are shown in Appendix Table C.1 below. Summarizing results shown in column (6), every metro has at least one eigenvalue greater than 1 and the median metro has 48% of its eigenvalues greater than 1. Ultimately, the reason that the system is not stable as measured by these eigenvalues is that households have very strong preferences over the racial composition of their neighbors. Restated, the racial composition of neighborhoods at the steady state implied by the current data is unstable because many households want to live in more segregated neighborhoods. This result is not merely a statement about the direction of racial preferences; it is more of a statement about the size of these preferences. To show this, we recompute eigenvalues of $\mathcal{G}$ holding $\delta^{\tau}_{\ell,m}$ fixed for all tracts ℓ, metros m, and types τ, but after multiplying all coefficients on race in utility, $\Theta_2^{\tau} = \{a_1^{\tau}, \ldots, a_5^{\tau}\}$, by 0.25 for all types and then by 0.125 for all types. By holding $\delta^{\tau}_{\ell,m}$ fixed, we preserve the relative desirability of all tracts in the baseline, so any changes to eigenvalues only reflect changes in the strength of preferences for race. The bottom line is that with these scaled-down preferences for race, stability for all metros vastly improves. Measured at the median metro, as shown in column (7) with the rescaling of Θ_2^{τ} by 0.25, 19.3% of a metro's eigenvalues are larger than 1, and with the rescaling of Θ_2^{τ} by 0.125, measured at the median metro 0.2% of a metro's eigenvalues are larger than 1, shown in column (8).

Appendix Table C.1

		Total			% Tracts Eigenvalues > 1		
Name	Pop (000s)	Tracts	% Black	% Hisp	Baseline	$0.25a_k^{\tau}$	$0.125a_k^{\tau}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Springfield, IL	201.4	55	9.6	0.9	48.2	34.5	0.0

[31] For each element i, we set Δ_i equal to 1.0×10^{-6}.

Name	Pop (000s)	Total Tracts	% Black	% Hisp	% Tracts Eigenvalues > 1 Baseline	$0.25a_k^\tau$	$0.125a_k^\tau$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Spartanburg, SC	253.8	51	21.0	2.7	49.0	45.1	4.9
Norwich, CT	259.1	62	6.5	5.2	51.6	12.9	0.0
Port St. Lucie, FL	319.4	60	11.8	8.0	57.5	25.0	0.8
Charleston, WV	309.6	76	5.0	0.6	42.8	9.9	0.7
Erie, PA	280.8	72	6.8	2.2	48.6	17.4	0.7
Eugene, OR	323.0	78	1.3	4.5	47.4	0.0	0.0
Montgomery, AL	346.5	82	40.7	1.1	45.1	34.1	18.3
Brownsville, TX	335.2	86	0.6	84.5	34.9	1.2	0.0
Salinas, CA	401.8	83	4.4	46.9	69.9	15.7	0.0
Utica, NY	299.9	92	5.1	2.6	45.7	4.9	0.0
Augusta, GA	499.7	95	35.6	2.4	47.4	41.6	15.3
Lansing, MI	447.7	117	9.0	4.7	48.7	31.2	0.0
Charleston, SC	549.0	117	31.0	2.4	48.7	44.4	17.5
Knoxville, TN	616.1	128	6.8	1.1	48.8	18.4	0.4
Greenville, SC	559.9	126	17.3	3.1	49.2	43.7	3.6
Worcester, MA	750.6	163	3.3	6.8	49.1	2.1	0.0
Youngstown, OH	603.0	168	11.2	1.7	44.0	20.8	0.0
Albany, NY	825.6	213	7.0	2.5	48.1	14.3	0.0
Dayton, OH	848.2	208	15.2	1.1	44.5	20.2	1.7
25th Percentile	309.6	76	5.1	1.7	45.7	12.9	0.0
Median	374.1	89	8.0	2.5	48.4	19.3	0.2
75th Percentile	559.9	126	15.2	5.2	49.0	34.1	3.6