A Unified Empirical Framework to Study Segregation

Gregorio Caetano and Vikram Maheshri*

February 1st, 2021

Abstract

We study the determinants of socioeconomic segregation in San Francisco Bay area neighborhoods from 1990 to 2004 with a novel empirical framework bridging the empirical literature on residential choice and the theoretical literature on neighborhood segregation. The former literature is based upon equilibrium empirical models of disaggregated choices, whereas the latter literature is concerned with the aggregate phenomenon of segregation, which is often studied theoretically in disequilibrium. Our framework explicitly allows for multiple equilibria and for data to be observed out of equilibrium. We identify the trajectory of convergence towards equilibrium allowing for sorting on the basis of the income and race compositions of neighborhoods, house prices, and unobserved amenities. These dynamics are explicitly mediated via two key potential frictions: incomplete information and moving costs. We also propose novel instrumental variables that exploit the logic of a dynamic choice model and can be constructed with no additional data requirements. Sorting based on unobserved neighborhood amenities is the most important factor generating segregation followed distantly by sorting based on the income and race of neighbors. Frictions, primarily moving costs, play a central role in keeping segregation in check, as they disproportionately mitigate discriminatory sorting on the basis of income and race.

*University of Georgia and University of Houston. All errors are our own. We gratefully acknowledge Patrick Bayer and Alvin Murphy for sharing their data and thank Milena Almagro, Josh Kinsler, Robert McMillan, Maisy Wong, anonymous referees and various conference and seminar participants for their helpful comments.

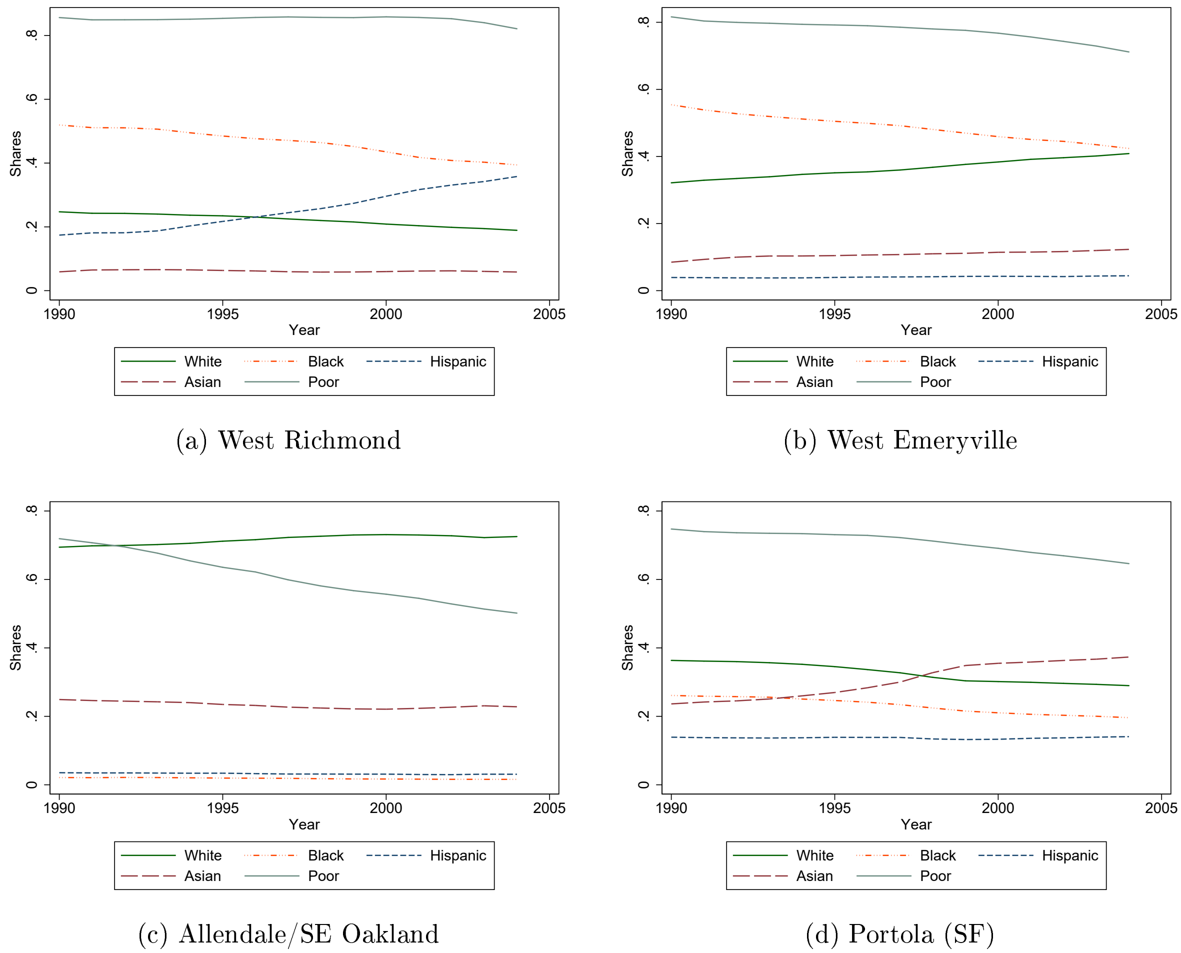
1 Introduction

There is a growing body of evidence that neighborhood segregation is an important engine of socioeconomic inequality.[1] Because similar households tend to have similar preferences, segregation is a natural outcome of residential sorting. Ever since Tiebout (1956), a rich theoretical and empirical literature has developed to study residential sorting and its consequences. Two strands of this literature have been particularly relevant to the study of segregation: empirical models of residential choice that have focused on studying the determinants of segregation in equilibrium (e.g., Bayer, McMillan and Rueben (2004*a*)), and disequilibrium models of segregation based on the seminal work of Schelling (1969) that have characterized segregation as an aggregate, dynamic phenomenon. In this paper, we introduce a unified framework to study segregation that bridges these two strands of the literature. We use this framework to study the determinants of racial and income segregation in the San Francisco Bay Area from 1990 to 2004.

In Figure 1, we show how the socioeconomic compositions of several neighborhoods in our sample have evolved over a fifteen year period. As suggested by these selected neighborhoods, we find rich heterogeneity in the trends of the race and income compositions across neighborhoods. These neighborhoods undergo substantial demographic changes in at least one socioeconomic dimension that often appear to be serially correlated. A common assumption in models of residential choice is that households are observed in what Bayer and Timmins (2005) define as a *sorting equilibrium*, i.e., in the absence of future amenity shocks, the demographic compositions of the neighborhoods will not change. Under this assumption, the trends shown in Figure 1 would be attributed to serially correlated, exogenous changes in the amenities of these neighborhoods. For instance, in the top left panel we would conclude that some West Richmond amenity that Hispanics disproportionately like systematically increased over the sample period in some manner outside of the model of residential choice. While this interpre-

[1]Residential segregation has been linked to a broad set of outcomes including educational attainment and labor market outcomes (Cutler, Glaeser and Vigdor (2008)), infant health (Mason et al. (2009)), friendship formation (Mouw and Entwisle (2006)), crime (Kling, Ludwig and Katz (2005)), intergenerational mobility and economic opportunity (Chetty et al. (2014)) and various measures of subjective well being (Ludwig et al. (2012)). Similarly, school segregation, a close cousin of neighborhood segregation, has been linked to lower educational attainment (Rivkin, Hanushek and Kain (2005)) and wider black-white achievement gaps (Card and Rothstein (2007)). In addition, school desegregation programs have been found to have increased black graduation rates (Guryan (2004)), college attendance and likelihood of arrest (Bergman (2016)).

Figure 1: Socioeconomic Composition of Selected Neighborhoods Over Time, 1990-2004



(a) West Richmond

(b) West Emeryville

(c) Allendale/SE Oakland

(d) Portola (SF)

tation may be appropriate when studying many aspects of residential sorting, it may be less appropriate when studying segregation.

Schelling (1969) has proposed an alternative interpretation of these trends that may be more plausible: the composition of a neighborhood may change even in the absence of shocks due to the presence of *discrimination*: households sort on the basis of the race or income of their neighbors. If, for instance, Hispanic households prefer Hispanic neighbors relative to non-Hispanic households, then an increase in the Hispanic share of a neighborhood could induce additional relative inflows of Hispanic households, which would in turn trigger further inflows of Hispanics in the future, and so on. This endogenous positive feedback loop mechanism would generate the observed serial correlation in socioeconomic composition we see in West Richmond even in the absence of other amenity shocks. The dynamics induced by such social interactions has led Schelling (1971) to suggest that neighborhoods are more likely to be observed adjusting along a trajectory to a sorting equilibrium rather than having reached that equilibrium already. While this interpretation is attractive in its simplicity – it offers a single, parsimonious explanation for the variety of demographic trends in all four of the neighborhoods above – disequilibrium models of segregation have remained mostly theoretical and have tended to focus on these endogenous responses to explain segregation while downplaying the potentially important roles of other factors.

This paper attempts to unify these two literatures in a new empirical framework to study segregation. We use it to study socioeconomic (race and income) segregation between eight types of households (rich and poor White, Black, Hispanic and Asian) in the San Francisco Bay Area from 1990-2004 using a recently constructed, high frequency data set on residential moves (Bayer et al. (2016)). We find that neighborhoods are far from equilibrium during our sample period. In fact, in the absence of any shocks, it would take many years for the vast majority of neighborhoods to converge to equilibrium. The effects of this slow adjustment would not be borne equally by all households. White and Black segregation would increase by 17% and 18% respectively, whereas Hispanic and Asian segregation would increase by 42% and 30% respectively; income segregation would increase by 60% but from such a low base that it is dominated by the increases in racial segregation.

Our framework allows us to analyze the relative importance of a variety of mechanisms that generate segregation – racial discrimination, income discrimination, sorting on the basis of other neighborhood amenities, differential responses to prices – and the frictions that mediate these mechanisms – moving costs, and incomplete information. We find that discrimination plays a very small role in shaping the levels of both race and income segregation. Rich households are less price-sensitive than poor households, but this difference also explains segregation levels little. Instead, sorting on the basis of other amenities matters the most. All of these segregation forces are substantially attenuated by frictions, especially moving costs. Moving costs are able to keep segregation in check by disproportionately mitigating the effects of discrimination. Absent moving costs, there would be a much larger amount of sorting across neighborhoods that would dramatically reshape the socioeconomic compositions of neighborhoods. This in turn would trigger further discriminatory sorting, thus creating a powerful feedback loop and establishing a more prominent role for discrimination as a determinant of segregation.

However, moving costs substantially attenuate this feedback loop, carving out a greater role for the other neighborhood amenities as determinants of segregation.

Our empirical framework combines a dynamic model of neighborhood choice with a simulation procedure that allows us to isolate specific determinants of segregation (and their interactions) by analyzing simulated counterfactuals. We summarize it here. First, households form expectations about the endogenous characteristics of neighborhoods (their race and income compositions, and their effects on prices) as well as other characteristics of neighborhoods. Based on these expectations, heterogeneous households decide if they should move, and if so, which neighborhood is best for them. Besides moving costs, we allow for an additional potential friction: at the time of their decision households may not know the exact levels of neighborhood amenities that they will encounter (including who their neighbors will be). We allow for this error in expectations because it may be particularly relevant when studying segregation. Indeed, this gives rise to the dynamics at the core of Schelling (1969), since households may not be able to coordinate their residential decisions. This coordination problem may be exacerbated in an environment where moving is costly since it is then more difficult to undo a decision that was made with expectations that went unrealized.

There are two key sets of parameters in this model that we must estimate for each of the eight socioeconomic groups of households: their moving costs, and their causal responses to different types of neighbors. Following Bayer et al. (2016), we identify moving costs from the decisions of households to move instead of staying in their current houses. We identify the causal responses of households to their neighbors through the use of novel instrumental variables (IVs) that are constructed with no additional data requirements. Briefly, our identification strategy relies on the assumption that information from the distant past (e.g., two years ago) does not directly affect valuations of neighborhoods today *conditional on valuations from the more recent past* (e.g., in the past year). We present empirical evidence that this assumption is likely to be satisfied in our context and show how this IV is consistent with our dynamic model with incomplete information and moving costs.

Given these estimated parameters, and our model, we simulate what would happen endogenously to the socioeconomic compositions of neighborhoods under various counterfactuals that include: different initial allocations of households across neighborhoods, different responses to neighbors (e.g., no race and/or income discrimination), different levels of moving costs, and different price sensitivities. For a simulation starting today, we explicitly model the fact that some households might decide to sort over the next month in response to changing expectations of the socioeconomic compositions of neighborhoods or their effects on prices, which may in turn trigger further moves two months from now, repeating this endogenous feedback loop indefinitely until reaching a new equilibrium. We simulate this entire dynamic re-sorting process to uncover the resulting trajectories of neighborhoods under these different counterfactuals. Comparing these trajectories across counterfactuals allow us to identify the relative roles of each factor in explaining segregation.

Relevant Literature

Our paper lies at the nexus of two distinct but related literatures on residential choice and segregation. We briefly review some of the most relevant studies.

Empirical Models of Residential Choice and Neighborhood Sorting

Because segregation is an outcome of neighborhood sorting, we build upon the prolific literature on the determinants of residential choice.[2] This literature is largely interested in estimating the marginal willingness to pay for neighborhood amenities. Three papers in this literature are particularly related to our study. Bayer, McMillan and Rueben (2004*a*) develop a framework to estimate horizontal models of neighborhood choice by building on insights from the empirical industrial organization literature (Berry (1994); Berry, Levinsohn and Pakes (1995)). This framework has been widely applied and extended in this literature (e.g., Bayer, McMillan and Rueben (2004*b*); Bayer, Ferreira and McMillan (2007); Bayer, Keohane and Timmins (2009); Ferreira (2010); Ringo (2013); Bayer et al. (2016); Caetano (2019)). They also discuss endogeneity that arises in the presence of an endogenous amenity such as the composition of neighbors. Bayer and Timmins (2005) study the existence and uniqueness of equilibrium in such sorting models with endogenous amenities, and Bayer and Timmins (2007) discuss estimation in empirical models like these and suggest an IV approach for identification based on the logic of a static model of neighborhood choice.

Our framework borrows several insights from these papers. As in this literature, we employ a discrete choice framework that enables us to study the relative importance of

[2] See, for example, Epple, Filimon and Romer (1984); Kiel and Zabel (1996); Epple and Sieg (1999); Epple, Romer and Sieg (2001, 2003); Bayer, McMillan and Rueben (2004*a*,*b*); Bayer and Timmins (2005, 2007); Bayer, Ferreira and McMillan (2007); Bayer et al. (2016); Caetano (2019). Kuminoff, Smith and Timmins (2013) provide a comprehensive review of the growing literature on neighborhood sorting.

racial and income composition versus other amenities in explaining the sorting patterns that lead to segregation. This also allows us to embed moving costs as an additional friction that prevents sorting. A key departure lies in our weakening of assumptions on households' expectations when residential decisions are made. We build a framework that is more agnostic about how expectations are formed. This is crucial for our purpose, as it renders our approach compatible with residential choices being observed out of equilibrium. Another related departure is that our framework suggests a different strategy to estimate a dynamic model of residential choice with moving costs. Although this is not the first paper to do so in the context of neighborhood choice (see, for example, Bayer et al. (2016) and Caetano (2019)), we show that many standard assumptions in dynamic demand estimation can be avoided when the goal is to study segregation (as opposed to estimating the value of amenities as is typical in these studies). Finally, the IV approach that we develop is novel, and it follows from the logic of a dynamic model of neighborhood choice. These IVs can be created with no additional data requirements and they can also be used to identify causal price responses.

Disequilibrium Models of Segregation

A largely theoretical literature based on the seminal Schelling model (Schelling (1969, 1971)) has sought to explore how segregation can arise and evolve when households care about the characteristics of their neighbors. In the Schelling model, heterogeneous agents select where to live by simple rules of thumb. Although this purely heuristic model is not explicitly based on the optimization of an objective, it generates valuable insight into a fundamental social force that may drive segregation: agents of different types react systematically differently to the composition of their neighbors. Schelling also makes explicit the role of some friction to ensure that neighborhoods gradually evolve toward an equilibrium state (e.g., myopia as in the original model).

Subsequent theoretical papers have embedded this intuition into a more standard economic framework (e.g., Becker and Murphy (2000); Bayer and Timmins (2005)), and there have been some recent attempts to estimate these models of segregation in reduced-form and structural contexts (e.g., Card, Mas and Rothstein (2008a); Banzhaf and Walsh (2013); Caetano and Maheshri (2017, 2020)). Banzhaf and Walsh (2013) discuss the role of exogenous amenities in generating segregation under no moving costs. Caetano and Maheshri (2017) and Caetano and Maheshri (2020) study school segregation in a framework that embeds the key insight of Schelling (1969). In this paper, we generalize and extend that framework in four directions. First, we analyze segregation along multiple dimensions simultaneously. Second, we make fewer assumptions on households' expectations, thus imposing fewer restrictions on the way that race and income compositions of neighborhoods may evolve. Third, we explicitly model realistic frictions such as moving costs, which motivates novel IVs. Lastly, we extend the framework to account for heterogeneous and endogenous responses to prices.

The rest of the paper proceeds as follows. In Section 2, we describe a data set of high frequency residential decisions in the San Francisco Bay Area. In Section 3 we present an empirical model of neighborhood segregation, articulate the specific assumptions required for identification, and discuss the estimation of this model. We present our baseline results in Section 4 and consider different counterfactuals in order to assess the importance of various determinants of segregation in Section 5. In Section 6, we extend our framework to explicitly incorporate prices and present additional results before concluding in Section 7.

2 Data

We use a monthly sample of all San Francisco Bay Area neighborhoods from January 1990 to November 2004. We define the San Francisco Bay Area as the six core counties (Alameda, Contra Costa, Marin, Santa Clara, San Francisco and San Mateo counties) that comprise the major cities of San Francisco, Oakland and San Jose and their surroundings, which we divide into neighborhoods by merging contiguous Census tracts until each resulting neighborhood contains approximately 10,000 households. Those neighborhoods with fewer than six annual home sales in our sample period are dropped leaving a total of 224 neighborhoods.

For each neighborhood in each month, we compute estimates of their race and income composition following the approach described in Bayer et al. (2016). Because high frequency data on the socioeconomic composition of neighborhoods is unavailable from standard sources (e.g., the Census) we must merge information from two main sources in order to construct these variables. The first source is Dataquick Information Services, a national real estate data service. Dataquick provides a detailed listing of *all* real estate transactions in the Bay Area including buyers' and sellers' names, buyer's mortgage information and property locations. The second source is a a dataset on mortgage applications published in accordance with the Home Mortgage Disclosure Act (HMDA) of 1975. Notably, HMDA data contains demographic information on mortgage applicants and the locations of properties that the applicants are buying. By linking these datasets on buyer's mortgage information and property locations, we can estimate how the demographics of neighborhoods change with each real estate transaction. With neighborhood-level estimates of the flows of households of different groups, we estimate the actual socioeconomic composition of each neighborhood by anchoring our flow estimates to the actual socioeconomic composition of each neighborhood per the 1990 US Census.[3]

We classify households into eight groups on the basis of four races (Whites, Blacks, Hispanics and Asians) and two income designations (rich or poor, depending on whether household income is greater than $50,000 in 1990 dollars)[4]. For expositional simplicity, we refer to Hispanics as a race rather than an ethnicity, and the other three racial groups include only non-Hispanic households.[5] For each race-income group g, neighborhood j and month t, we observe the total number of homeowners, the total numbers of households who moved into a new house, and the total number of households who stayed in the same home since last month.[6] We also observe the total number of households of each group who chose to exit the Bay Area homeownership market in each month.[7]

We summarize our data in Table 1. The majority of homeowners in the Bay Area

[3]Bayer et al. (2016) report the results of multiple diagnostic tests that ensure the validity of this estimation procedure.

[4]We obtain the race and income of the original stock of households as of 1990 from the 1990 Census. From 1990 onwards, all changes in the income of the neighborhood are measured based on income data from HMDA deflated to 1990 levels. We chose an income threshold of $50,000 because it resulted in the most balance of rich and poor among all available thresholds in the 1990 Census.

[5]We are unable to observe populations at the race-ethnicity-income group-tract level in the 1990 Census. Instead, we are able to observe populations at the race-income group-tract level, at the ethnicity-income group-tract level, and at the race-ethnicity-tract level. As such, our raw counts of rich and poor Whites, Blacks and Asians in each neighborhood include Hispanics. To address this, we reweight each group uniformly across neighborhoods to ensure that the number of rich Whites plus the number of poor Whites is equal to the number of non-Hispanic Whites (and do the same for Blacks and Asians), and we uniformly reweight each group to ensure that the number of rich Hispanics plus the number of poor Hispanics is equal to the total number of Hispanics. Our results are effectively unchanged if we assume all Hispanics to be White and correct the population numbers accordingly.

[6]Households who move between houses within the same neighborhood counted as inflows (but not stayers).

[7]They are the households who are observed to move out of some neighborhood in $t-1$ but not observed to move into any neighborhood in t.

are White, although there are sizable Asian and Hispanic populations as well. Roughly 47% of homeowners in the Bay Area are classified as rich, though this share is much smaller for Blacks and Hispanics. The socioeconomic compositions of neighborhoods also change over time in our sample as reflected in monthly inflow rates ranging from 0.1% for poor Whites to 0.7% for rich Asians.

Table 1: Summary Statistics

	White		Black		Hispanic		Asian	
	Rich	Poor	Rich	Poor	Rich	Poor	Rich	Poor
Share of Homeowners	0.38	0.33	0.02	0.04	0.04	0.06	0.08	0.05
	(0.24)	(0.22)	(0.02)	(0.04)	(0.04)	(0.06)	(0.07)	(0.05)
Average Num. of Homeowners	2,196	1,879	108	244	217	340	471	315
	(2,089)	(1,161)	(145)	(370)	(280)	(330)	(342)	(342)
Average Monthly Inflows	8.71	2.36	0.38	0.30	1.26	0.90	3.58	1.13
	(10.57)	(3.36)	(0.92)	(0.78)	(2.71)	(1.73)	(6.60)	(2.02)
Average Monthly Stayers	2,187	1,877	108	243	216	340	467	314
	(2,081)	(1,160)	(145)	(370)	(278)	(329)	(702)	(341)
Dissimilarity Index	0.29	0.19	0.41	0.57	0.27	0.34	0.36	0.33
Num. of Observations				39,872				

Note: Each observation is a neighborhood-month from January 1990 to November 2004. Poor households have an income of less than \$50,000 in 1990 dollars. Standard deviations are presented in parentheses.

The high variance in the average number of homeowners of each group reflects substantial cross-sectional heterogeneity in the socioeconomic composition of neighborhoods, i.e., segregation. We calculate the dissimilarity index of for each of the eight groups defined by race and income, and summarize it in Table 1.[8] We choose this widely used measure of segregation because it is easy to interpret. For instance, a rich

[8]If N_{gj} is the total number of group g households residing in neighborhood j, then the dissimilarity

White dissimilarity index of 0.29 indicates that 29% of rich Whites would have to be relocated (holding all other households' locations fixed) in order to generate a uniform distribution of them across all Bay Area neighborhoods. The index ranges from zero to one, and a higher value means that households of a given socioeconomic group are more concentrated in certain neighborhoods. Blacks are the most concentrated racial group, followed distantly by Asians, Hispanics and Whites. While rich Whites and Asians tend to be more concentrated than their poor counterparts, the opposite is true for rich Blacks and Hispanics.

3 Empirical Framework

We start by providing an overview of our framework without getting bogged down in the details of identification and estimation, which are provided in the rest of this section. A city is divided into J neighborhoods, each of which are populated by households of G different demographic groups. Let N_{gjt} represent the number of group g households who reside in neighborhood j in period t. In every period, each neighborhood possesses a multidimensional endogenous amenity: the socioeconomic (race and income) composition of their residents denoted as the vector of shares $\boldsymbol{s_{jt}}$, where s_{gjt} is an element of this vector representing the share of group g:

$$s_{gjt} = \frac{N_{gjt}}{\sum_{g'} N_{g'jt}} \tag{1}$$

(Hereafter, we refer to all vectors and matrices in bold type.) The compositions of all neighborhoods in the city can be represented by the state matrix $\boldsymbol{s_t}$ whose jth column is $\boldsymbol{s_{jt}}$. At the beginning of each period, households form expectations of their value of residing in each neighborhood and then choose where to reside.

We start from a generic function for the aggregated group g demand for neighborhood j in period t:

$$N_{gjt} = f_{gj}\left(\boldsymbol{s_t^e}; \boldsymbol{\beta_g}, \phi_g\right) \tag{2}$$

where f_{gj} is a function unique to each group-neighborhood combination. We denote

index for group g households is defined as $\frac{1}{2}\sum_j \left|\frac{N_{gj}}{\sum_k N_{gk}} - \frac{N_j - N_{gj}}{\sum_k (N_k - N_{gk})}\right|$ where $N_j = \sum_g N_{gj}$. Note that a group may corresponds to a race-income combination (e.g., rich Whites), a race (e.g., rich Whites plus poor Whites) or an income level (e.g., poor households of all races).

$\boldsymbol{s_t^e} = \mathbb{E}_t\left[\boldsymbol{s_t}\right]$ as the expectation of $\boldsymbol{s_t}$ that is formed by households *just before they make their decision in t.* This expected quantity stands in contrast to $\boldsymbol{s_t}$, which is the *actual* composition *after* everyone made their decision in t. The parameter vector $\boldsymbol{\beta_g}$ represents the marginal effect of $\boldsymbol{s_t^e}$ on demand in the absence of moving costs, and the parameter vector ϕ_g represents the moving cost that group g households would incur if they moved out of the house they occupied in $t-1$.

Although N_{gjt} and $\boldsymbol{s_t}$ are observed, $\boldsymbol{s_t^e}$ is not, so it is infeasible to estimate $\boldsymbol{\beta_g}$ and ϕ_g directly from equation (2). To circumvent this issue, we use the actual, observed vector $\boldsymbol{s_t}$ as a proxy for $\boldsymbol{s_t^e}$, yielding

$$N_{gjt} = f_{gj}\left(\boldsymbol{s_t}; \boldsymbol{\beta_g}, \phi_g\right) + \underbrace{f_{gj}\left(\boldsymbol{s_t^e}; \boldsymbol{\beta_g}, \phi_g\right) - f_{gj}\left(\boldsymbol{s_t}; \boldsymbol{\beta_g}, \phi_g\right)}_{\text{error}_{gjt}} \tag{3}$$

With appropriate restrictions on f_{gj}, the parameters $\boldsymbol{\beta_g}$ and ϕ_g can be estimated in a dynamic discrete choice model that we develop in Section 3.1.

Given estimates of $\hat{\boldsymbol{\beta}}_{\boldsymbol{g}}$ and $\hat{\phi}_g$, we can analyze how the compositions of neighborhoods might evolve under different counterfactual values of $\boldsymbol{s_t^e}$. At $\boldsymbol{s_t^e} = \tilde{\boldsymbol{s}}$, the counterfactual demand of group g households for neighborhood j is equal to

$$N_{gjt}\left(\tilde{\boldsymbol{s}}\right) = f_{gj}\left(\tilde{\boldsymbol{s}}; \hat{\boldsymbol{\beta}}_{\boldsymbol{g}}, \hat{\phi}_g\right) \tag{4}$$

from which we obtain

$$s_{gjt}\left(\tilde{\boldsymbol{s}}\right) = \frac{N_{gjt}\left(\tilde{\boldsymbol{s}}\right)}{\sum_{g'} N_{g'jt}\left(\tilde{\boldsymbol{s}}\right)} \tag{5}$$

Calculating equation (5) for each group g yields the matrix-valued function $\boldsymbol{s_t}\left(\tilde{\boldsymbol{s}}\right)$, whose jth column is $\boldsymbol{s_{jt}}\left(\tilde{\boldsymbol{s}}\right)$ with generic element $s_{gjt}\left(\tilde{\boldsymbol{s}}\right)$. By considering any counterfactual value of $\tilde{\boldsymbol{s}}$, we can identify $\boldsymbol{s_t}\left(\cdot\right)$ by simulation. This function defines a dynamic system that completely characterizes the counterfactual evolution of neighborhood-level demographics (and thus segregation) from any initial state in the absence of future shocks, starting from t. If we repeatedly evaluate $\boldsymbol{s_t}\left(\cdot\right)$ starting from $\tilde{\boldsymbol{s}}$, we can construct the *simulated trajectory* $\mathbb{T}_{\boldsymbol{t}}\left(\tilde{\boldsymbol{s}}\right)$, which is a sequence whose first element is $\mathbb{T}_t^0(\tilde{\boldsymbol{s}}) = \tilde{\boldsymbol{s}}$, and whose subsequent elements can be defined recursively as $\mathbb{T}_t^\tau(\tilde{\boldsymbol{s}}) = \boldsymbol{s}_t\left(\mathbb{T}_t^{\tau-1}(\tilde{\boldsymbol{s}})\right)$.

Building on Bayer and Timmins (2005), we define a *sorting equilibrium* as a state $\boldsymbol{s}^\star$ which is the limit point of this trajectory.

Definition 1. *Sorting Equilibrium.* State $\boldsymbol{s}^\star$ is a *sorting equilibrium* if $\mathbb{T}_{\boldsymbol{t}}\left(\tilde{\boldsymbol{s}}\right)$ converges to $s^\star$ for some $\tilde{\boldsymbol{s}}$.

Throughout the paper, we appeal to this definition when we refer to "equilibrium". We say that neighborhoods in t are observed out of equilibrium if $\mathbb{T}_{\boldsymbol{t}}(\boldsymbol{s_t})$ does not converge to $\boldsymbol{s_t}$.

Remark 1. A key difference between our approach and other empirical approaches to study segregation is that ours allows for the demographic compositions of neighborhoods to be observed out of equilibrium. This is directly related to assumptions on households' expectations. Note that $\boldsymbol{s_t}(\boldsymbol{s^e_t}) = \boldsymbol{s_t}$ by construction since observed choices in t are made when $\tilde{\boldsymbol{s}} = \boldsymbol{s^e_t}$. Thus, assuming households perfectly forecast the compositions (i.e., $\boldsymbol{s^e_t} = \boldsymbol{s_t}$) implies $\boldsymbol{s_t}(\boldsymbol{s_t}) = \boldsymbol{s_t}$, i.e., it implies that data are observed in equilibrium. Similarly, small deviations from this assumption (e.g., households share a common information set apart from zero-mean private information) will also imply that data are assumed to be observed in equilibrium. More generally, the trajectory of convergence towards equilibrium is likely affected by expectations, so it is crucial that we avoid strong assumptions on the formation of households' expectations if we want to study the dynamics of segregation.

3.1 Identification

We now formalize a model of neighborhood choice that gives rise to equation (2) and impose restrictions that allow for identification and feasible estimation of the function $f_{gj}(\cdot; \boldsymbol{\beta_g}, \phi_g)$ with available data. At any period t, household i faces the dynamic optimization problem

$$\max_{j_{i\tau} \in \mathbb{J}} \mathbb{E}_t \left[\sum_{\tau=t}^{\mathcal{T}} \delta^{\tau-t} \cdot u(j_{i\tau}, \boldsymbol{b_{i\tau}}) \,|j_{it}, \boldsymbol{b_{it}} \right], \tag{6}$$

where $j_{i\tau}$ and $\boldsymbol{b_{i\tau}}$ are the choice and state variables of household i in period τ, respectively, $\mathbb{J}$ is each household's choice set, $u(\cdot)$ is their flow indirect utility function, $\mathcal{T}$ is their time horizon, and δ is their inter-temporal discount factor.

We define the value function as $V(\boldsymbol{b_{it}}) = \max_{j \in \mathbb{J}} v(j, \boldsymbol{b_{it}})$, where the choice-specific value function is written as

$$v(j, \boldsymbol{b_{it}}) = u(j, \boldsymbol{b_{it}}) + \int \delta \cdot V(\boldsymbol{b_{it+1}})\, dF_b(\boldsymbol{b_{it+1}} \,|j, \boldsymbol{b_{it}}). \tag{7}$$

$F_b(\boldsymbol{b_{it+1}}\,|j, \boldsymbol{b_{it}})$ is the expected distribution of the state variable in $t+1$ conditional on the choice and the state variable from t.

Assumption 1. (Additive Separability, Logit Error) *We can decompose* $\boldsymbol{b_{it}} = (\boldsymbol{x_{it}}, \boldsymbol{\epsilon_{it}})$ *so that* $u(j, \boldsymbol{b_{it}}) = u(j, \boldsymbol{x_{it}}) + \epsilon_{ijt}$*, where* ϵ_{ijt} *(the* j*-th element of* $\boldsymbol{\epsilon_{it}}$*) is i.i.d. extreme value type I.*

Assumption 1 implies

$$v(j, \boldsymbol{b_{it}}) = \underbrace{u(j, \boldsymbol{x_{it}}) + \int \delta \cdot V(\boldsymbol{b_{it+1}})\, dF_b(\boldsymbol{b_{it+1}}\,|j, \boldsymbol{x_{it}})}_{v(j, \boldsymbol{x_{it}})} + \epsilon_{ijt}, \tag{8}$$

At the beginning of period t, households observe the state variable $\boldsymbol{b_{it}}$ and choose (a) whether or not to move, and upon deciding to move, they then choose (b) an option in $\mathbb{J} = \{0, \ldots, J\}$. Options $j \in \{1, ..., J\}$ correspond to residing in neighborhood j. Option $j = 0$ corresponds to the outside option of residing outside of the city.[9] Following Bayer et al. (2016), we simplify notation and index the option of staying in the same home in one of the J neighborhoods as option $J + 1$. We also denote v_{gjt} as the average of $v(j, \boldsymbol{x_{it}})$ across all households from group g.

Following Bayer et al. (2016), we impose the following restriction on the moving cost parameters ϕ_g to ensure that the identification and estimation of cumulative utilities is feasible with available data:

Assumption 2. *Group* g *households who decide to move incur a fixed moving cost* ϕ_g *irrespective of their neighborhoods of origin* (j_{it-1}) *and destination* (j_{it})*.*

Under this assumption, we can specify $v(j, \boldsymbol{x_{it}})$ from equation (8) as

$$v(j, \boldsymbol{x_{it}}) = \mathbf{1}(j \in \{0, \ldots, J\}) \cdot (v_{gjt} - \phi_g) + \mathbf{1}(j = J + 1) \cdot v_{gjt} \tag{9}$$

where $\mathbf{1}(\cdot)$ is the indicator function. While the first term of equation (9) represents the cumulative utility of choosing to move and then choosing either one of the neighborhoods or the outside option, the second term refers to the cumulative utility of choosing to stay in the same home as last period. v_{gjt} is specified as

$$v_{gjt} = \boldsymbol{\beta_g'} \boldsymbol{s_{jt}^e} + \xi_{gjt} \tag{10}$$

[9] As in Bayer et al. (2016), we only observe data on homeowners, so in our application, $j = 0$ also corresponds to the outside option of renting within the city.

This specification is without loss of generality, as equation (10) simply projects v_{gjt} separately by group into a component dependent on the expected socioeconomic composition of the neighborhood and a remainder.[10]

We define the vector of state variables $\boldsymbol{x_{it}} = (j_{it-1}, \boldsymbol{s^e_t}, \boldsymbol{\xi_{g_i t}})$, where g_i is the demographic group to which i belongs, and $\boldsymbol{\xi_{g_i t}}$ is the vector whose j-th element is $\xi_{g_i jt}$.[11] We now make the following assumption, which implicitly restricts what is included in $\boldsymbol{\xi_{g_i t}}$ (see Remark 3):

Assumption 3. *$Cov\left(\xi_{gjt}, \boldsymbol{s_{jt-T}} | inflows_{\boldsymbol{jt-1}}, ..., inflows_{\boldsymbol{jt-T'}}\right) = 0$ for some $T > T' \geq 1$.*

Under Assumption 3, we can identify $\boldsymbol{\beta}_g$ using $\boldsymbol{s_{jt-T}}$ as an instrumental variable (IV) for $\boldsymbol{s^e_{jt}}$ once we control for a flexible function of inflows in intermediate periods between t and $t-T$.

Intuitively, our identification strategy is based on the idea that if neighborhood amenities or expectations change after some period in the distant past (in $t-T$ or before), some households will no longer find themselves residing in their most desired neighborhood. When they first moved to the neighborhood in the distant past, it was their optimal choice given their expectations (since they maximized their utility in equation (6)). However, moving costs may have "locked" them into that neighborhood. Because expectations about neighborhood amenities continue to evolve after households move (either because amenities actually change or otherwise), this lock-in can accumulate *mismatch* between each household's current neighborhood and its newly most desired neighborhood. Only when this mismatch exceeds moving costs do households re-sort to their most desired neighborhood, in turn resetting their mismatch to zero. In a context where moving costs are sufficiently high and amenities or expectations change sufficiently over time, a great deal of mismatch may have accumulated at any moment in the data, which our IV strategy exploits.

Therefore, our identification strategy exploits an asymmetry in equation (10): while

[10] Note that we do not assume that $\boldsymbol{s^e_{jt}}$ enters linearly in equation (10). Rather, any non-linearity will be embedded in ξ_{gjt}, which will have implications for Assumption 3. This allows us to interpret $\boldsymbol{\beta'_g s^e_{jt}}$ as the best linear approximation of the (potentially non-linear) relationship between $\boldsymbol{s^e_{jt}}$ and v_{gjt}. In the estimation section we show how in practice by adding flexible controls in this equation we can weaken this validity assumption as much as data allows, and we also discuss a variety of robustness checks we conduct in Footnote 26.

[11] Note that different households of the same group are allowed to differ from each other only via their previous choice, j_{it-1}, and $\boldsymbol{\epsilon_{it}}$. This restriction can be weakened if additional data is available (e.g., Berry, Levinsohn and Pakes (1995)).

$\boldsymbol{s^e_{jt}}$ is (the expectation of) a *stock* variable, v_{gjt} is only a *flow* variable.[12] To illustrate this, we consider the simplest case with $T = 2$. As before, let the superscript "e" refer to expectations of the corresponding variable taken by households just before their decisions are made in t. We can decompose a generic scalar element of $\boldsymbol{s^e_{jt}}$, s^e_{gjt}, into expected inflows and expected stayers as follows:

$$
\begin{aligned}
s^e_{gjt} =& \frac{N^e_{gjt}}{\sum_{g'} N^e_{g'jt}} \\
=& \frac{\text{inflows}^e_{gjt} + \text{stayers}^e_{gjt}}{\sum_{g'} \left(\text{inflows}^e_{g'jt} + \text{stayers}^e_{g'jt}\right)} && (11)\\
=& \frac{\text{inflows}^e_{gjt} + \pi^e_{gjt} \cdot N_{gjt-1}}{\sum_{g'} \left(\text{inflows}^e_{g'jt} + \pi^e_{g'jt} \cdot N_{g'jt-1}\right)} && (12)\\
=& \frac{\text{inflows}^e_{gjt} + \pi^e_{gjt} \cdot (\text{inflows}_{gjt-1} + \pi_{gjt-1} \cdot N_{gjt-2})}{\sum_{g'} \left(\text{inflows}^e_{g'jt} + \pi^e_{g'jt} \cdot (\text{inflows}_{g'jt-1} + \pi_{g'jt-1} \cdot N_{g'jt-2})\right)} && (13)
\end{aligned}
$$

Equation (11) follows from the accounting identity $N_{gjt} = \text{inflows}_{gjt} + \text{stayers}_{gjt}$. Equation (12) follows from substituting $\pi^e_{gjt} \cdot N_{gjt-1}$ for stayers^e_{gjt} into the previous equation, where π^e_{gjt} is the expected probability (as of t) that a group g household stays in the same home in neighborhood j from $t-1$ to t. Finally, equation (13) follows from substituting $\text{inflows}_{gjt-1} + \pi_{gjt-1} \cdot N_{gjt-2}$ for N_{gjt-1} into the previous equation.[13]

Note that the IV, $s_{gjt-2} = \frac{N_{gjt-2}}{\sum_{g'} N_{g'jt-2}}$, and the endogenous variable of interest, s^e_{gjt}, are potentially correlated, since both are functions of $\boldsymbol{N_{jt-2}} = (N_{1jt-2}, ..., N_{Gjt-2})'$. Based on equation (13), s_{gjt-2} may be correlated to s^e_{gjt} via three paths:

1. $\pi^e_{g'jt} \cdot \pi_{g'jt-1} \neq 0$, i.e., some residents of j in $t-2$ are expected to remain there in t even if what originally led them to reside there no longer affects future inflows ($\text{Cov}\left(\text{inflows}^e_{g'jt}, s_{gjt-2}\right) = \text{Cov}\left(\pi^e_{g'jt} \cdot \text{inflows}_{g'jt-1}, s_{gjt-2}\right) = 0$).

2. $\text{Cov}\left(\pi^e_{g'jt} \cdot \text{inflows}_{g'jt-1}, s_{gjt-2}\right) \neq 0$, i.e., some residents who moved into j in $t-1$ for reasons that are correlated to s_{gjt-2} are expected to remain in t.

3. $\text{Cov}\left(\text{inflows}^e_{g'jt}, s_{gjt-2}\right) \neq 0$, i.e., some residents are expected to move into j in t for reasons that are correlated to s_{gjt-2}.

[12] The term "flow" here is in contrast to "stock", so it means something different from the term "flow" used elsewhere in the paper.

[13] For simplicity, the superscript "e" is dropped in $t-1$ or before, since as of t households might already know these values from the past (although that is not assumed, see Remark 3).

We would like to exploit the first path for causal identification of $\boldsymbol{\beta}_g$ since it does not affect households who move in t and is therefore uncorrelated to ξ_{gjt}. However, it is confounded by the second and third paths. By flexibly controlling for $\text{inflows}_{g'jt-1}$ for all g', we eliminate the second path. Finally, once the second path is eliminated, the second path is also eliminated by Assumption 3. In words, our exclusion restriction can be restated as "no information that was relevant to decision-making in $t-2$ or before (i.e., correlated to $\boldsymbol{s_{jt-2}}$) and irrelevant to inflows in $t-1$ (i.e., uncorrelated to $\text{inflows}_{g'jt-1}$ for all g') is relevant to inflows in t (i.e, correlated to $\boldsymbol{\xi_{jt}}$)." For a given choice of T', increasing T weakens the three paths of correlation (and weakens Assumption 3) and hence represents a tradeoff of instrument relevance (first path) for validity (eliminating the second and third paths).[14]

Interpretation of $\boldsymbol{\beta}$

First, note that to the extent that households discriminate on the basis of their neighbors' socioeconomic status, this will be included in $\boldsymbol{\beta}$. Because $\boldsymbol{\beta}$ captures the response to the socioeconomic composition of neighborhoods, it contains both pure socioeconomic animus (or affinity) and statistical discrimination by socioeconomic status, and these are not separately identified.

Second, since $\boldsymbol{\beta}$ captures a full reduced-form response to changes in the socioeconomic composition of neighborhoods, we do not separately identify whether it is mediated through the associated change in the flow value of a neighborhood, the continuation value of a choice, or both. To see this, consider equation (8), and let u_{gjt} and CV_{gjt} be, respectively, the averages of $u\left(j, \boldsymbol{x_{it}}\right)$ and $\int \delta \cdot V\left(\boldsymbol{b_{it+1}}\right) dF_b\left(\boldsymbol{b_{it+1}} | j, \boldsymbol{x_{it}}\right)$ across all group g households. Then we can write $v_{gjt} = u_{gjt} + CV_{gjt}$, where u_{gjt} represents the flow utility component and CV_{gjt} is the continuation value of households who choose neighborhood j in t. For each g and g', we identify $\beta_{g,g'} = \frac{\partial v_{gjt}}{\partial s^e_{g'jt}} = \frac{\partial u_{gjt}}{\partial s^e_{g'jt}} + \frac{\partial CV_{gjt}}{\partial s^e_{g'jt}}$, i.e. the total marginal effect of an expected increase in g' share on the group g valuation of that neighborhood.

We opt to estimate this reduced-form effect because it allows us to study many aspects of segregation without imposing additional assumptions that are needed for this decomposition (see Remark 2).[15] As Manski (2004) argues, choice data alone is

[14] We perform a series of robustness checks that compare estimates for $T = 13, ..., 36$ for $T' = 12$; we find that the relevance of the IV, which is derived from the first path of correlation, remains strong as T grows, yet the estimates do not change. See Figures 13 and 14 in the appendix.

[15] In Section 6, we isolate one of these channels, the price channel, in order to consider a specific

insufficient to separately identify expectations and preferences. For instance, if we observed poor Whites leaving a neighborhood because of an increase in the poor Black share, choice data alone would not permit us to infer that they responded to a prejudice against poor Blacks (a preference) as opposed to a signal that the neighborhood would become less desirable to them in the future for some other reason (an expectation), or both.[16] This is particularly important when studying segregation since assumptions on expectations directly impact simulated trajectories (Remark 1).

Finally, note that although moving costs $\boldsymbol{\phi}$ are explicitly separate from $\boldsymbol{\beta}$ in the model, this only applies to moves in period t. That is, the expected costs of future moves (in response to unforeseen changes in neighborhood characteristics) are not contained in $\boldsymbol{\phi}$, but they do impact CV_{gjt}, hence by the argument above, they are loaded in $\boldsymbol{\beta}$. As a result, $\boldsymbol{\beta}$ also contains a component related to errors in households' expectations; to the extent that households are unable to perfectly predict the future socioeconomic compositions of neighborhoods, their responses to the current compositions will be affected since they will take this into account when deciding to move today.

Remark 2. Standard dynamic discrete choice approaches often impose the conditional independence assumption $F_b\left(b_{it+1}|j, x_{it}\right) = F_x\left(x_{it+1}|j, x_{it}\right) \cdot G_\epsilon\left(\epsilon_{it+1}\right)$, where $G_\epsilon\left(\epsilon_{it+1}\right)$ refers to the extreme value type I distribution.[17] This restricts how individuals expect (as of t) neighborhoods to be in the future depending on their choice and their state variable. As discussed in Remark 1, to study segregation it is important to remain as agnostic as possible about households' expectations. By avoiding this assumption, we allow for the expected transitions of households of different demographic groups to differ from one another. This implies, for instance, that the time horizon, $\mathcal{T}$, and the inter-temporal discount factor, δ, may vary across groups, and neither needs to be observed or identified.

Remark 3. The restriction on $\boldsymbol{\xi_{git}}$ in Assumption 3 implies that inflows in t do not use past information (from $t-T$ or before) in a more sophisticated manner than inflows of some group in $t-1, ..., t-T'$. This is a restriction on the relative level of sophistication, not on the absolute level, so it is consistent with many formulations of expectations,

counterfactual.

[16]While this would prevent us from identifying, say, poor White households' willingness to pay to avoid residing close to poor Black neighbors, it would not restrict us from analyzing how poor Whites sort into or out of a neighborhood in response to an increase in the poor Black share since this is fundamentally related to households' choices and not their preferences *per se*.

[17]See, e.g., Aguirregabiria and Mira (2010) for a great survey of the literature. To facilitate comparison with standard approaches, we use their notation whenever possible.

ranging from the narrowly myopic households of Schelling (1969) to highly sophisticated households. For instance, consider households with rational expectations who use their information set in the best way possible (their forecast errors are orthogonal to their information set). Let those making decisions in t form their expectations with information from the last τ periods, while those making decisions in $t - T'$ form their expectation with information from the last τ' periods. Then the assumption implies $\tau \leq \tau' + T'$, so households in t are allowed to be somewhat more sophisticated than those in $t - T'$ (i.e., $\tau > \tau'$ is allowed).[18] In particular, $(\text{inflows}_{\boldsymbol{jt-1}}, \ldots, \text{inflows}_{\boldsymbol{jt-T'}})$ is not required to be in the information set of group g households in t since we use only the variation in $\boldsymbol{s_{gt-T}}$ that is orthogonal to it in order to identify $\boldsymbol{\beta_g}$.

3.2 Estimation and Simulation

Our empirical approach unfolds in three stages: we first estimate v_{gjt} and ϕ_g for all g, j and t (stage 1) and then we estimate the causal effect of $\boldsymbol{s^e_{jt}}$ on v_{gjt} (stage 2). Finally, we use these estimates to simulate the evolution of the demographic compositions of neighborhoods under different counterfactuals (stage 3).

Stage 1: Estimation of v_{gjt} and ϕ_g

This stage follows closely from Bayer et al. (2016). First, we use the choices of only those who moved in period t to estimate the cumulative utilities v_{gjt}. Having decided to move, household i solves the following optimization problem:

$$\max_{j \in \{0,\ldots,J\}} v_{g_i jt} - \phi_{g_i} + \epsilon_{ijt} \tag{14}$$

Following Assumption 1, the choice-specific probabilities are

[18] Although at first our identification strategy might look similar to strategies used in the production function literature, such as the "proxy variable" literature (e.g., Olley and Pakes (1996)) and the dynamic panel literature (e.g., Arellano and Bond (1991)), there are important differences. In our approach we exploit an identifying assumption that relates the information set of one decision maker (households making their decision in t) with the information set of another decision maker (households of several groups making their decisions in the past). This only requires us to be restrictive with the information set of the household in t *relative* to the information set of the households in the past; we do note have to impose *absolute* restrictions on their information set. In these other contexts, identification exploits absolute restrictions in the information set of a given decision maker (e.g., firms) at the time of their decision. See Ackerberg (2020) for an illuminating discussion of these two literatures in the context of restrictions on information sets at the time decisions are made.

$$
\begin{aligned}
P\left(j_{it}=j \,|\, j\notin\{J+1\}, j_{it-1}\right) &= \frac{\exp\left(v_{g_ijt}-\phi_g\right)}{\sum_{j'=0}^{J}\exp\left(v_{g_ij't}-\phi_g\right)} \\
&= \frac{\exp\left(v_{g_ijt}\right)}{\sum_{j'=0}^{J}\exp\left(v_{g_ij't}\right)}
\end{aligned}
\tag{15}
$$

Because moving costs are assumed to not vary by the neighborhood of origin or destination (Assumption 2), they cancel out. Following Berry (1994), we estimate $\hat{v}_{gjt}$ for $j\in\{0,\ldots J\}$ as

$$
\hat{v}_{gjt}=\log\left(\text{inflows}_{gjt}\right)-\log\left(\text{inflows}_{g0t}\right). \tag{16}
$$

Next, we consider the choice of whether to stay in the same home to identify the moving cost parameter ϕ_g. For household i who resided in j last period, the probability of choosing option $J+1$ (i.e., not moving) is

$$
\begin{aligned}
P\left(j_{it}=J+1 \mid j_{it-1}=j\right) &= P\left(v_{g_ijt}+\epsilon_{iJ+1t}>v_{g_ij't}-\phi_{g_i}+\epsilon_{ij't}\,,\ \forall j' |\, j_{it-1}=j\right) \\
&= \frac{\exp\left(v_{g_ijt}\right)}{\sum_{j'=0}^{J}\exp\left(v_{g_ij't}-\phi_{g_i}\right)+\exp\left(v_{g_ijt}\right)}
\end{aligned}
\tag{17}
$$

where the first line must hold for all $j'=0,\ldots,J$, and the second line follows from the logit formula (Assumption 1). The data analog to $P\left(j_{it}=J+1\mid j_{it-1}=j\right)$ is simply $\frac{\text{stayers}_{g_ijt}}{N_{g_ijt-1}}$, or the proportion of group g_i households residing in neighborhood j in $t-1$ who decided to stay in the same home in the following period. Hence, equation (17) yields the J moment restrictions

$$
h_j\left(\phi_g;\hat{\boldsymbol{v}}_{\boldsymbol{gt}}\right)=\frac{\text{stayers}_{gjt}}{N_{gjt-1}}-\frac{\exp\left(\hat{v}_{gjt}\right)}{\sum_{j'=0}^{J}\exp\left(\hat{v}_{gj't}-\phi_g\right)+\exp\left(\hat{v}_{gjt}\right)} \tag{18}
$$

By plugging in our estimates of $\hat{v}_{gjt}$ from equation (16) into equation (17), we can estimate ϕ_g by GMM using moment conditions (18).

Stage 2: Estimation of $\boldsymbol{\beta_g}$

Following equation 10, we decompose $\hat{v}_{gjt}$, which was estimated in the first stage, as

$$\hat{v}_{gjt} = \boldsymbol{\beta_g'} \boldsymbol{s_{jt}} + \boldsymbol{\Lambda_g} \left(\boldsymbol{\hat{v}_{jt-1}}, ..., \boldsymbol{\hat{v}_{jt-T'}} \right) + \gamma_{gt} + \tilde{\xi}_{gjt} \tag{19}$$

where $\boldsymbol{\Lambda_g}(\cdot)$ is a flexible function[19], γ_{gt} is a group-period fixed effect, and $\tilde{\xi}_{gjt}$ is an error term that includes all remaining unobserved determinants of $\hat{v}_{gjt}$. The parameters of interest, $\boldsymbol{\beta_g}$, represent the causal effects of $\boldsymbol{s_{jt}}$ on $\hat{v}_{gjt}$. We use $\boldsymbol{s_{jt-T}}$ as an Instrumental Variable (IV) for $\boldsymbol{s_{jt}}$ to estimate $\boldsymbol{\beta_g}$ via Two-Stage Least Squares.

Because we do not observe v_{gjt} or $\boldsymbol{s^e_{jt}}$, we use $\hat{v}_{gjt}$ and $\boldsymbol{s_{jt}}$, respectively, as proxies for them. Subtracting equation (10) from equation (19) and rearranging yields the following error decomposition:

$$\tilde{\xi}_{gjt} = \left(\xi_{gjt} - \boldsymbol{\Lambda_g} \left(\boldsymbol{\hat{v}_{jt-1}}, \ldots, \boldsymbol{\hat{v}_{jt-T'}} \right) \right) + \boldsymbol{\beta}'_g \left(\boldsymbol{s^e_{jt}} - \boldsymbol{s_{jt}} \right) + \left(\hat{v}_{gjt} - v_{gjt} \right) \tag{20}$$

The first term corresponds to the determinants of households' cumulative utilities that are due to amenities other than $\boldsymbol{s^e_{jt}}$ (ξ_{gjt} from equation (10)) and that are orthogonal to $(\boldsymbol{\hat{v}_{jt-1}}, ..., \boldsymbol{\hat{v}_{jt-T'}})$. Assumption 3 implies that the IV is uncorrelated to this term.[20] The second term corresponds to forecast errors in households' expectations. Note that v_{gjt} reflects choices made with the same information set that was used to form $\boldsymbol{s^e_{jt}}$. Hence, forecast errors $\left(\boldsymbol{s^e_{jt}} - \boldsymbol{s_{jt}} \right)$ cannot affect decisions in t.[21] Finally, the third term corresponds to any error in estimation of households' cumulative utilities that arose from the first stage. Assumptions 1 and 2 imply that the first stage estimates are consistent, thus $\hat{v}_{gjt} - v_{gjt}$ will be uncorrelated to our IV.

[19] In practice we specify $\boldsymbol{\Lambda_g}$ as cubic B-splines of $\hat{v}_{g'jt-\tau}$ for each g' and τ. There are four knots for each element. Note that each of the coefficients of these variables are allowed to vary by g.

[20] Equation (16) implies a direct relationship between $\text{inflows}_{g'jt-\tau}$ and $\hat{v}_{g'jt-\tau}$: $\hat{v}_{g'jt-\tau} = \log \text{inflows}_{g'jt-\tau} - \log \text{inflows}_{g'0t-\tau}$. Since $-\log \text{inflows}_{g'0t-\tau}$ does not vary across neighborhoods, it is absorbed by γ_{gt}. Thus, Assumption 3 implies that controlling *flexibly* for $\boldsymbol{v_{jt-1}}, ..., \boldsymbol{v_{jt-T'}}$ and γ_{gt} in equation (19) yields consistent estimates of $\boldsymbol{\beta_g}$. In practice, controlling flexibly for $\text{inflows}_{\boldsymbol{jt-\tau}} = (\text{inflows}_{1jt-\tau}, ..., \text{inflows}_{Gjt-\tau})$ instead yields estimates that are statistically indistinguishable from our main estimates.

[21] More specifically, any variation in $\boldsymbol{s_{jt}}$ that affects v_{gjt} must do so through $\boldsymbol{s^e_{jt}}$. It follows that any relevant IV of $\boldsymbol{s_{jt}}$ would affect v_{gjt} only through $\boldsymbol{s^e_{jt}}$ and not through $\boldsymbol{s^e_{jt}} - \boldsymbol{s_{jt}}$.

Stage 3: Identifying Sorting Equilibria by Simulation

Once we obtain estimates of $\hat{v}_{gjt}$, $\hat{\phi}_g$ and $\hat{\boldsymbol{\beta}}_{\boldsymbol{g}}$, we can identify by simulation how the demographic composition of each neighborhood will evolve from any initial state in the absence of external shocks. Consider, for instance, period t as our starting point. We denote the population distribution of the entire city with group-specific population vectors $\boldsymbol{N_{gt}} = (N_{g1t}, \ldots, N_{gJt})$, which imply share vectors $\boldsymbol{s_{gt}} = (s_{g1t}, \ldots s_{gJt})$ and a demographic composition matrix $\boldsymbol{s_t}$ with generic element $\boldsymbol{s_{gt}}$.

For any given counterfactual matrix of expected compositions of neighborhoods $\tilde{\boldsymbol{s}} = (\tilde{\boldsymbol{s}}_{\boldsymbol{1}}, ..., \tilde{\boldsymbol{s}}_{\boldsymbol{J}})$, we write the counterfactual expected valuation of neighborhood j by group g households as

$$v_{gjt}(\tilde{\boldsymbol{s}}) = \hat{v}_{gjt} + \hat{\boldsymbol{\beta}}'_{\boldsymbol{g}} (\tilde{\boldsymbol{s}}_{\boldsymbol{j}} - \boldsymbol{s_{jt}}) \tag{21}$$

where $\hat{v}_{gjt}$ is the estimated cumulative value of the neighborhood, whereas $v_{gjt}(\tilde{\boldsymbol{s}})$ is computed under the counterfactual group composition $\tilde{\boldsymbol{s}}$. This equation simply removes the part of the valuation due to $\boldsymbol{s_{jt}}$ and adds the part due to $\tilde{\boldsymbol{s}}_{\boldsymbol{j}}$. Note that equation (21) implies that ξ_{gjt} is kept constant in the simulation.

We simultaneously simulate counterfactual group-specific valuations for all neighborhoods from $\tilde{\boldsymbol{s}}$ as

$$\begin{aligned} \tilde{N}_{gjt}(\tilde{\boldsymbol{s}}) \;&=\; N_{gjt-1}\left(\frac{\exp\left(v_{gjt}(\tilde{\boldsymbol{s}})\right)}{\exp\left(-\hat{\phi}_g\right) + \sum_{j'=1}^{J}\exp\left(v_{gj't}(\tilde{\boldsymbol{s}}) - \hat{\phi}_g\right) + \exp\left(v_{gjt}(\tilde{\boldsymbol{s}})\right)}\right) + \\ &\quad + \sum_{k=1}^{J} N_{gkt-1}\left(\frac{\exp\left(v_{gjt}(\tilde{\boldsymbol{s}}) - \hat{\phi}_g\right)}{\exp\left(-\hat{\phi}_g\right) + \sum_{j'=1}^{J}\exp\left(v_{gj't}(\tilde{\boldsymbol{s}}) - \hat{\phi}_g\right) + \exp\left(v_{gkt}(\tilde{\boldsymbol{s}})\right)}\right) \end{aligned} \tag{22}$$

The first term on the right-hand side of equation (22) corresponds to the simulated number of households who resided in neighborhood j in $t-1$ and remained in their house, incurring no moving costs. The second term represents the simulated number of households who resided in neighborhood k in $t-1$ and then moved to neighborhood j next period (households with $k = j$ moved within neighborhood j). Because our simulation holds fixed all factors that affect households' propensity to choose the outside option, $v_{g0t}(\tilde{\boldsymbol{s}})$ is constant, so $v_{g0t}(\tilde{\boldsymbol{s}}) = v_{g0t} = 0$, where the second equality follows from the normalization in period t.[22]

[22]Net inflows into the Bay Area from $t-1$ to t can be accounted for by simply re-weighting the

Once we obtain $\tilde{N}_{gjt}(\tilde{\boldsymbol{s}})$, we can calculate

$$s_{gjt}(\tilde{\boldsymbol{s}}) = \frac{\tilde{N}_{gjt}(\tilde{\boldsymbol{s}})}{\sum_{g'} \tilde{N}_{g'jt}(\tilde{\boldsymbol{s}})} \tag{23}$$

Note that $s_{gjt}(\tilde{\boldsymbol{s}})$ is the $(g,j)^{\text{th}}$ element of the function $\boldsymbol{s_t}(\tilde{\boldsymbol{s}})$, which characterizes the evolution of the demographic compositions of all neighborhoods in the absence of shocks. By repeatedly evaluating $\boldsymbol{s_t}(\cdot)$ starting from $\tilde{\boldsymbol{s}}$, we can construct the simulated trajectory $\mathbb{T}_{\boldsymbol{t}}(\tilde{\boldsymbol{s}})$ using equations (21), (22) and (23). Given a sufficiently fine grid, a tolerance μ and a time threshold $\bar{t}$, the state $\boldsymbol{s}^{\star}$ for which $\|\mathbb{T}_{\boldsymbol{t}}^{\boldsymbol{\tau}}(\tilde{\boldsymbol{s}}) - \boldsymbol{s}^{\star}\| < \mu$ for all $\tau > \bar{t}$ is interpreted as a sorting equilibrium (Definition 1). We can identify all equilibria this way by conducting a grid search of all possible counterfactual states $\tilde{\boldsymbol{s}}$ and computing the simulated trajectory $\mathbb{T}_{\boldsymbol{t}}(\tilde{\boldsymbol{s}})$ for each counterfactual.[23]

4 Empirical Results

Our empirical analysis covers eight socioeconomic groups – all combinations of four races and two income groups – each of whom are allowed to respond heterogeneously to unobserved amenities as well as to four endogenous amenities – the shares of Blacks, Hispanics, and Asians (relative to Whites) and the share of the poor (relative to the rich).[24]

In Table 2, we present estimates of the causal responses to the socioeconomic compositions of neighborhoods ($\boldsymbol{\beta_g}$), and the moving costs (ϕ_g) for households of each group. The endogenous amenities $\boldsymbol{s_{jt}}$ are instrumented by $\boldsymbol{s_{jt-13}}$, and $\boldsymbol{\Lambda_g}$ is specified as a cubic B-spline of each element of $v_{g'jt-\tau}$ for all g' and $\tau = 1, ..., 12$.[25] We choose $T' = 12$

simulated $\tilde{N}_{gjt}(\tilde{\boldsymbol{s}})$ to ensure that $\sum_{j=1}^{J} \tilde{N}_{gjt}(\tilde{\boldsymbol{s}}) = \sum_{j=1}^{J} N_{gjt}$ for all t. In practice, this does not affect our results.

[23] One might be concerned that the linear specification in equation (19) does not allow for flexibility in the simulated trajectories to equilibrium. In previous work (Caetano and Maheshri (2017)), we show that a simple linear specification already provides substantial flexibility even in a far more restrictive context.

[24] We lack sufficient data to precisely estimate $\boldsymbol{\beta}$ if we allowed each of the eight groups to respond to race and income in an unrestricted, non-separable way (i.e., estimate 8x7 instead of 8x4 estimates of $\beta_{g,g'}$). With more data this could be implemented.

[25] For each of the 8 groups g' and 12 lags τ, there are 4 knots of each element $\hat{v}_{g'jt-\tau}$, which yields a total of $8 \cdot 12 \cdot 4 = 384$ control variables. We allow the coefficients of each of these control variables to vary by g.

in order to control for seasonal effects.

Since White (poor) share is the omitted race (income) amenity, the responses $\beta_{g,g'}$ are interpreted as the response of group g to a marginal increase in $s^e_{g'jt}$ relative to a marginal increase in the share of White (rich) neighbors. We find that households of each group respond positively to neighbors of the same race and to neighbors of the same income. Poor White households are more responsive to poor neighbors than rich White households, but this pattern is reversed for minority households. Hispanics respond most positively to neighbors of their own race, followed by Asians and Blacks. Interestingly, not all responses are reciprocated: e.g., Blacks respond negatively to Hispanics, but Hispanics show little response to Blacks. There is also heterogeneity in the interaction between race and income. e.g., rich Asians respond less intensely than poor Asians to same-race neighbors, but the opposite is true for Blacks and Hispanics. Altogether, these heterogeneous responses may give rise to complex dynamics.

Although we estimate mostly statistically significant discriminatory responses, the key takeaway from this table is that these responses are small in comparison to our estimates of moving costs that are an order of magnitude larger and precisely estimated. This suggests that substantial amenity mismatch may accumulate since many households may be locked into a neighborhood that is no longer their most preferred neighborhood. Although our estimates of moving costs are generally statistically different from each other, they are similar in magnitude across all socioeconomic groups (the maximum variation in these moving costs is less than 10% of the estimates).

In the Appendix, we present OLS estimates of $\boldsymbol{\beta}$ (Table 5). These are much larger in magnitude than our IV estimates since there are many confounding reasons why similar households would choose similar neighborhoods (e.g., they tend to value other amenities more similarly), all of which biasing the OLS estimates upward in magnitude. The OLS bias is most pronounced for the within-group parameter estimates, as expected. We also report estimates of $\boldsymbol{\beta}$ for different values of T (the period corresponding to our IV) in Figures 13 and 14. Larger values of T weaken Assumption 3 resulting in an IV that is more likely to be valid. We find that all 12 parameter estimates change very little for $T = 13, \ldots, 36$.[26]

[26]We also performed two other types of robustness checks in order to assess whether controlling flexibly for $\boldsymbol{v_{jt-1}}, ..., \boldsymbol{v_{jt-12}}$ was sufficient to absorb confounders: (1) We re-estimated $\boldsymbol{\beta}$ under different specifications of $\boldsymbol{\Lambda_g(\cdot)}$ (linear rather than cubic B-spline, different number of knots, and using inflows$_{g'jt-\tau}$ instead of $v_{g'jt-\tau}$) and obtained similar estimates for all flexible specifications; (2) we further controlled for average neighborhood prices $P_{jt-1}, ..., P_{jt-12}$ and obtained similar estimates of all $\boldsymbol{\beta}$ coefficients. We also attempted to increase the value of T', but this check turned out to be unin-

Table 2: Responses to the Race and Income Compositions of Neighborhoods ($\boldsymbol{\beta}$) and Moving Costs ($\boldsymbol{\phi}$)

	White		Black		Hispanic		Asian	
	Rich	Poor	Rich	Poor	Rich	Poor	Rich	Poor
Responses to:								
Black Share	-2.83***	-2.25***	3.45***	2.41***	0.49	-0.12	-1.22***	-0.83**
	(0.40)	(0.45)	(0.43)	(0.42)	(0.40)	(0.43)	(0.43)	(0.41)
Hispanic Share	-4.57***	-1.18	-1.42**	-3.22***	11.41***	8.85***	-0.25	0.02
	(0.67)	(0.91)	(0.66)	(0.69)	(0.81)	(0.81)	(0.77)	(0.79)
Asian Share	-0.49	-3.97***	0.34	-1.17***	-0.81	-1.97***	5.05***	7.37***
	(0.48)	(0.63)	(0.52)	(0.49)	(0.61)	(0.62)	(0.67)	(0.68)
Poor Share	-0.69***	3.60***	-2.35***	0.44	-1.67***	0.17	-2.94***	1.42***
	(0.24)	(0.42)	(0.37)	(0.29)	(0.39)	(0.34)	(0.37)	(0.40)
Moving Costs	28.65***	28.70***	26.96***	28.18***	27.50***	28.71***	28.03***	28.30***
	(0.02)	(0.02)	(0.05)	(0.03)	(0.03)	(0.01)	(0.02)	(0.01)
R^2	0.79							
Num. of Observations	147,840							

Notes: This specification includes group-month fixed effects and control variables $\boldsymbol{\Lambda_g}\left(\boldsymbol{\hat{v}_{jt-1}}, ..., \boldsymbol{\hat{v}_{jt-12}}\right)$ (see Footnote 25) as well as $s_{g'jt-13}$ for each of the four socioeconomic shares as instrumental variables. White is the omitted racial share and rich is the omitted income share. All standard errors clustered by group-month. * - 90% significance, ** - 95% significance, *** - 99% significance. The p-values for both the Cragg-Donald and the Kleinbergen-Paap weak identification tests are less than 0.01, which implies a strong first stage.

With these estimates, we simulate how the socioeconomic compositions of neighborhoods would evolve keeping $\xi_{g'jt}$ constant for all g' (i.e., in the absence of future external

formative as it yielded imprecise estimates since each additional lag increased the numbers of controls in $\boldsymbol{\Lambda_g}(\cdot)$ dramatically.

shocks). We focus on the counterfactual $\tilde{\boldsymbol{s}} = \boldsymbol{s_t}$ where t refers to the final month of our sample, November 2004, and conduct the simulation until convergence. In Figure 2, we present a graph of the number of neighborhoods that experience at least 1, 2, 5 or 10 simulated moves that change their socioeconomic composition. If, for instance, a rich White homeowner simply left a neighborhood, that would count as one change (one outflow). If instead they were replaced by another rich White homeowner, that would count as zero changes. If they were replaced by a homeowner of a different race or income level, that would count as two changes (one outflow plus one inflow). We describe neighborhoods experiencing such changes to their socioeconomic compositions as "in flux."

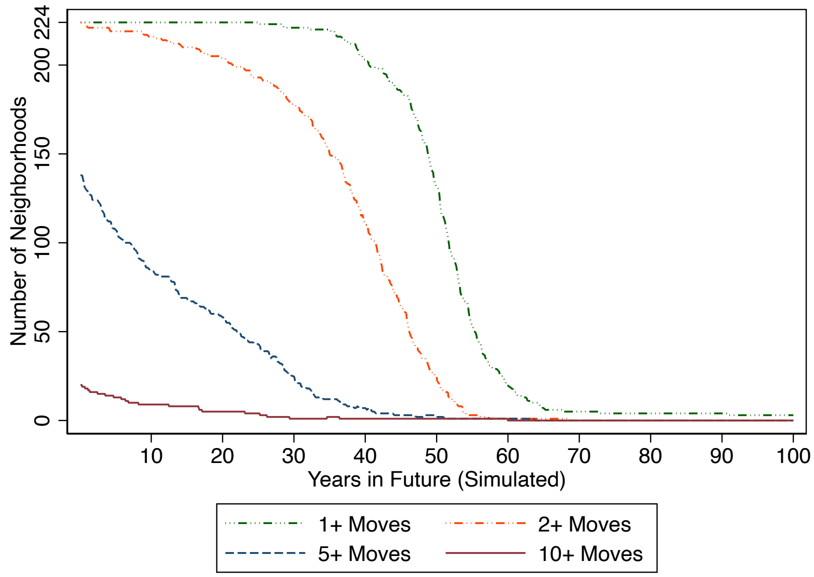
Initially, and for decades to follow, nearly all neighborhoods are in flux. From this, we conclude that neighborhoods are actually not observed to be in equilibrium. Despite substantial moving costs, the amenities of the neighborhoods where households reside are sufficiently unattractive to some households at present that most neighborhoods experience turnover. Over time, changes in the socioeconomic compositions of these neighborhoods feedback and also spill over to other neighborhoods, which in turn changes their relative attractiveness to homeowners of all socioeconomic groups. This process is slow, as it takes 60-70 years for the Bay Area to approximate sorting equilibrium.[27]

The outcome of this pattern of sorting is a change in the levels of segregation in the Bay Area. In Figure 3, we present the long run change in the dissimilarity index for each race (pooling income groups) and for each income group (pooling races) across all Bay Area neighborhoods.

Over the course of adjustment, all races experience modest increases in segregation. White households experience the smallest increase in segregation in both absolute and relative (19%) terms. Black households start off more segregated than all other races and remain so throughout the simulation. Hispanic homeowners experience the largest absolute and relative (42%) increases in segregation, followed by Asians, who experience a 31% increase in segregation. Homeowners are least segregated by income initially and remain so throughout the simulation despite the largest relative increase (62%).

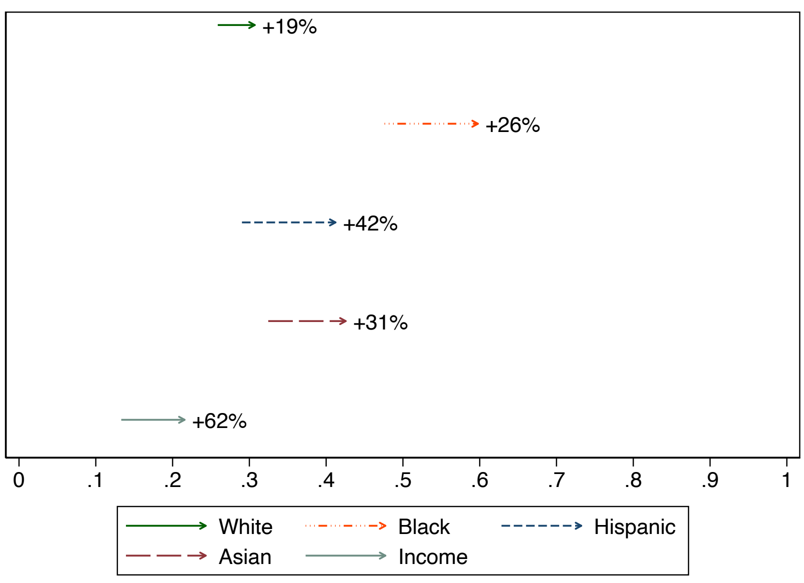
[27]It takes 153 years for all Bay Area neighborhoods to experience no moves (see Appendix Figure 15).

Figure 2: Number of Neighborhoods In Flux (Simulated)



Notes: Figure shows the number of neighborhoods with at least one, two, five or ten moves that change their socioeconomic composition (out of a total of 224 neighborhoods). Simulation begins in November 2004.

Figure 3: Equilibrium Changes in Race and Income Segregation (Simulated)



Notes: The arrows represent the changes in simulated Dissimilarity Indices for households of each race and income from November 2004 to sorting equilibrium in the absence of exogenous shocks. Numbers correspond to the relative change in dissimilarity. A Black dissimilarity index of, say, 0.60, means that 60% of Black homeowners would have to be relocated in order to generate an equal distribution of Blacks across all Bay Area neighborhoods.

5 Determinants of Long-Run Segregation

In this section we study the roles of discrimination, moving costs, incomplete information and the initial allocations of households in explaining the long-run changes in segregation that we found in Section 4. We weigh the importance of these determinants by leveraging the various moving parts of our framework to simulate several relevant counterfactuals. This ensures that we allow for complex sorting patterns to emerge that would otherwise be difficult to predict but are nonetheless integral to the dynamic process of segregation. Indeed, the discriminatory responses that we estimate may not necessarily increase segregation as one may expect. For instance, rich White homeowners fleeing a neighborhood that is becoming more Black will, all else constant, increase not only the Black share of neighbors, but also the Hispanic and Asian shares of neighbors. That in turn may lead to further inflows of not only Blacks, but also Hispanics and Asians.[28] The complexity of this sorting pattern grows over time not only because all groups continue to respond endogenously to each of these changes in a given neighborhood, but also because they respond to concomitant changes in other neighborhoods.

5.1 The Roles of Discriminatory Responses: Race and Income

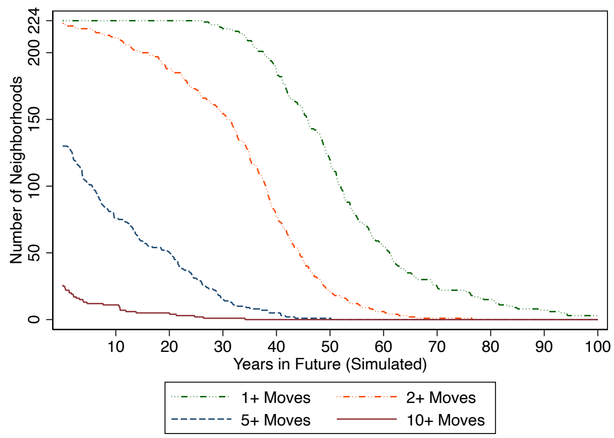
Our estimates of $\hat{\boldsymbol{\beta}}$ reveal systematic discriminatory responses for homeowners of all socioeconomic groups. To isolate their roles in explaining the patterns of segregation dynamics presented in Figures 2-3, we consider a series of counterfactuals in which households are either "race-blind", i.e., indifferent to the racial composition of their neighbors, "income-blind", i.e., indifferent to the income composition of their neighbors, or both race- and income-blind. As shown in Figure 4, discriminatory responses have little qualitative effect on segregation dynamics, though they do slightly slow down the process of arriving at sorting equilibrium.

We present the simulated increase in segregation under each of these counterfactuals in Figure 5. Baseline effects from Figure 3 are reproduced in light gray. This

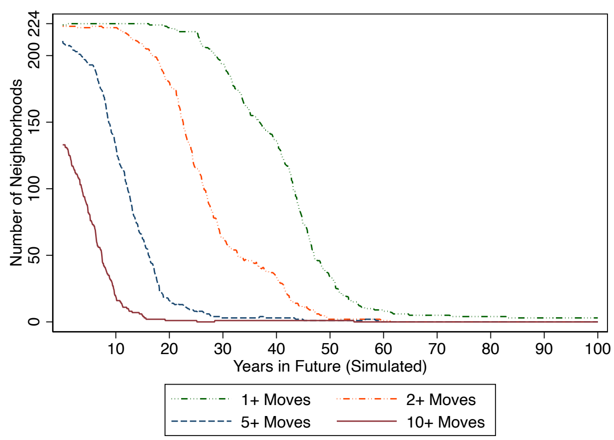
[28]As shown in Table 2, households tend to respond more positively to an increase in the share of same-race households than negatively to an increase in the shares of other races.

Figure 4: Number of Neighborhoods In Flux - No Discrimination (Simulated)

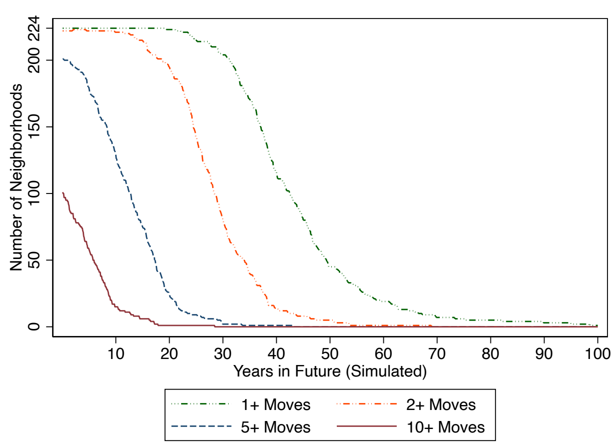
(a) No Racial Discrimination



(b) No Income Discrimination

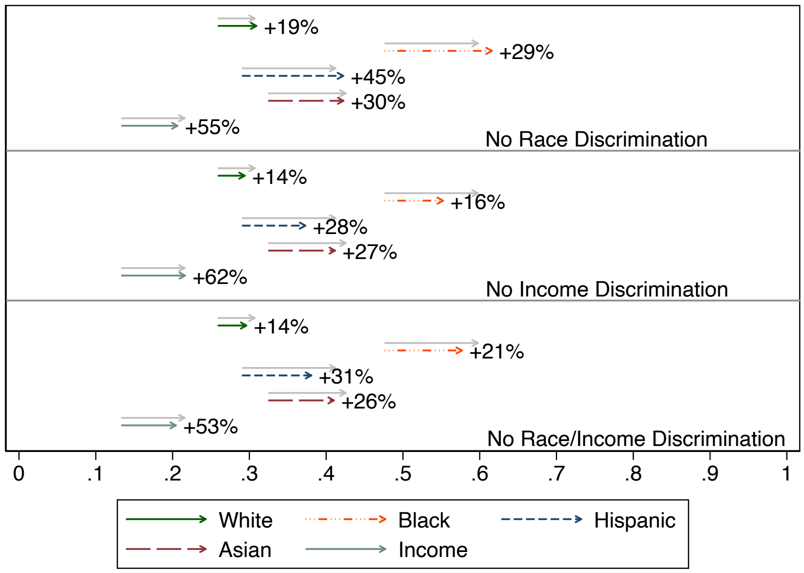


(c) No Racial or Income Discrimination



Notes: Each panel shows the number of neighborhoods with at least one, two, five or ten moves that change their socioeconomic composition (out of a total of 224 neighborhoods) under a different counterfactual. Simulation begins in November 2004.

Figure 5: Equilibrium Changes in Race and Income Segregation - No Discrimination (Simulated)



Notes: For each panel representing a different counterfactual, the arrows represent changes in simulated Dissimilarity Indices for households of each race and income from November 2004 to sorting equilibrium in the absence of exogenous shocks. Numbers in parentheses correspond to the relative change in dissimilarity. Gray arrows correspond to the baseline changes shown in Figure 3. A Black dissimilarity index of, say, 0.60, means that 60% of Black homeowners would have to be relocated in order to generate an equal distribution of Blacks across all Bay Area neighborhoods.

makes it apparent that removing discrimination, either by race-blinding households, income-blinding households, or both, has little impact on segregation (with the exception of Hispanic segregation, for which income discrimination does seem to matter more). Together, discriminatory responses to the race and income of neighbors account for at most 10% of the baseline long-run increase in segregation. In contrast, the bottom panel shows that sorting on the basis of amenities other than the socioeconomic compositions of neighborhoods explains most of the long-run changes in segregation.

5.2 The Role of Moving Costs

The gradual declines of Figures 4 and 4 suggest that moving costs play an important role in the dynamics of segregation. To explore this further, we consider a counterfactual in which all homeowners enjoy a one-time amnesty of zero moving-costs at the beginning of the simulation. As shown in the first panel of Figure 6, the Bay Area converges to a sorting equilibrium instantaneously. In the first period, the lack of moving costs allows households to eliminate their mismatch (per their expectation). However, this does not imply that there is no mismatch in further periods, as errors in households' expectations may lead them to reside in neighborhoods that are suboptimal. Nevertheless, this mismatch is insufficient to overcome the substantial moving costs which are restored in future periods. This happens because each element of $\boldsymbol{\beta}$ is small in magnitude relative to moving costs. In order to illustrate this point, consider a different counterfactual where we eliminate moving costs both today and in every future period by permanently setting $\boldsymbol{\phi} = \mathbf{0}$. We do not consider this counterfactual to be particularly sensible because households would have to be repeatedly surprised by the future elimination of moving costs every time it occurs.[29] However, this exercise is valuable as it allows us to gauge the role of incomplete information, which is hidden in our context but would perhaps play a more important role in a context with much lower moving costs or much higher discrimination than the one we encounter in our sample. The dynamics of this second counterfactual are shown in the second panel of Figure 6. While convergence is still much faster than in the baseline case with moving costs, it is not instantaneous, owing to the fact that errors in prediction would still trigger further moves (which would remain costless in this scenario), leading to the feedback loop discussed in Schelling (1969).[30]

We explore the interaction between moving costs and segregation in Figure 7, which is the analog to Figure 3 when moving costs are eliminated at the beginning of the simulation, but resume in the future in whichever manner households took them into account when making the choices we observed in the data. This counterfactual in the top

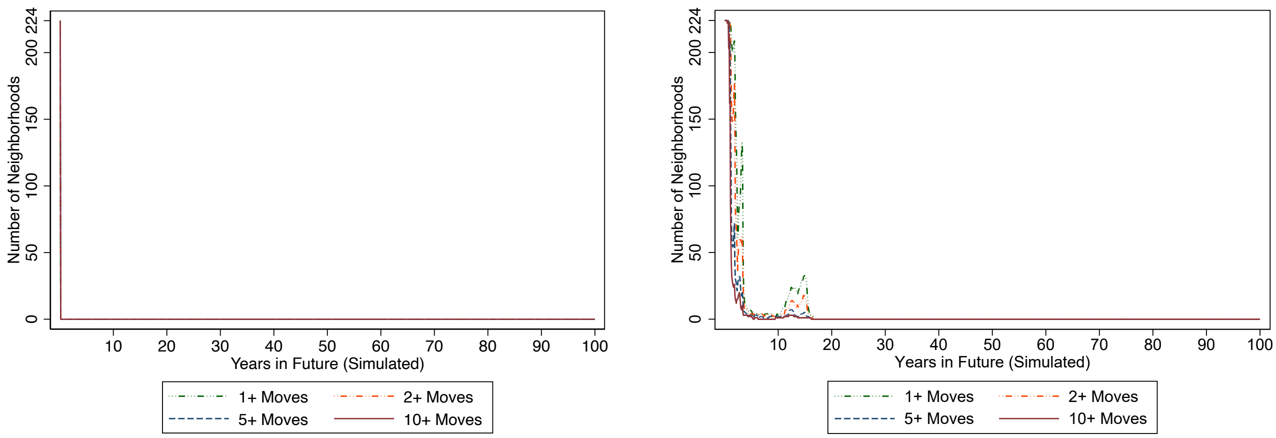
[29]This "surprise" must occur because $\boldsymbol{\beta}$ also contains the marginal effects of $\boldsymbol{s_{jt}}$ on the continuation values and was estimated using data generated in a world with expectations of non-zero future moving costs (see "Interpretation of $\boldsymbol{\beta}$" in Section 3.1). Note that we would encounter a related issue if we instead decided to separately identify the flow utility and the continuation value components of $\boldsymbol{\beta}$. Doing so would require us to assume households are forward looking and anticipate future moving costs in a very specific way, but if this assumption was invalid, it would lead to misestimation of the simulated trajectories. For instance, in practice households may discount the future differently depending on their socioeconomic group.

[30]If we also set $\boldsymbol{\beta} = \mathbf{0}$ in the second counterfactual, then we obtain instantaneous convergence to the equilibrium again. This is expected, since in this case we remove not only all scope for discrimination, we also ensure that future valuations of neighborhoods are not different from expected valuations of those neighborhoods as of t because of errors in prediction of their socioeconomic compositions.

Figure 6: Number of Neighborhoods In Flux - No Moving Costs (Simulated)

(a) No Initial Moving Costs

(b) No Moving Costs Ever

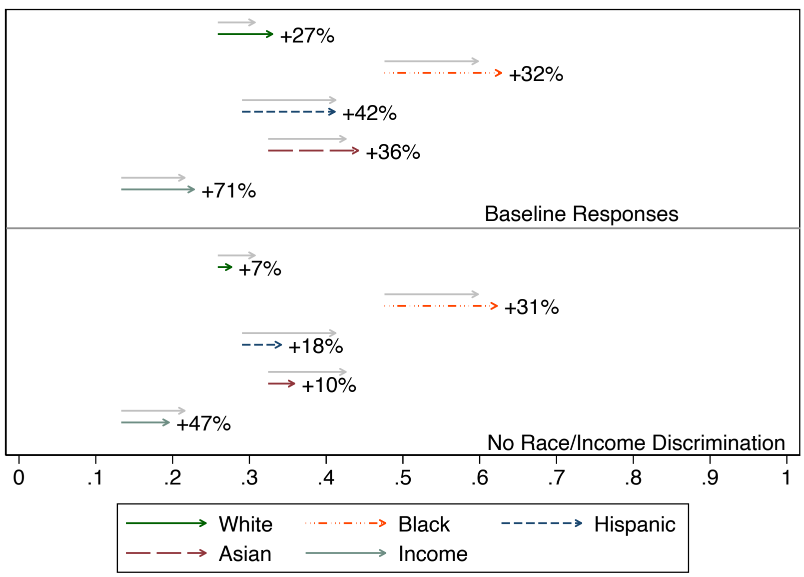


Notes: Each panel shows the number of neighborhoods with at least one, two, five or ten moves that change their socioeconomic composition (out of a total of 224 neighborhoods) under a different counterfactual. Simulation begins in November 2004.

panel (which maintains baseline discriminatory responses) leads to higher segregation levels across the board relative to the baseline effects shown in the gray arrows. When we further shut off discriminatory responses (lower panel), the resulting increases in segregation are much less pronounced.

A comparison of the two panels of Figure 7 reveals that discrimination plays an important role in a world without moving costs. For instance, in such a world, discriminatory sorting would be responsible for a 24% increase in Hispanic segregation (42-18=24). In reality, discrimination is much less important than sorting towards other amenities (as shown in the bottom panel of Figure 5); because moving is costly, only a few households sort at a time, which reduces the scope for socioeconomic changes to trigger future moves. In Figure 16 in the appendix we show analogous results for the counterfactual with no moving costs ever. As expected, segregation would increase substantially in that counterfactual (first panel) but much less under no discriminatory responses (second panel), highlighting a greater role for the endogenous feedback loop. Because moving costs after the first adjustment are still zero in this counterfactual, the feedback loop is more intense, and the importance of discrimination relative to sorting on the basis of other amenities increases. Thus, we conclude that frictions, especially moving costs, disproportionately mitigate the role of discrimination on segregation, leaving more space for other amenities to play a larger role.

Figure 7: Equilibrium Changes in Race and Income Segregation - No Initial Moving Costs (Simulated)

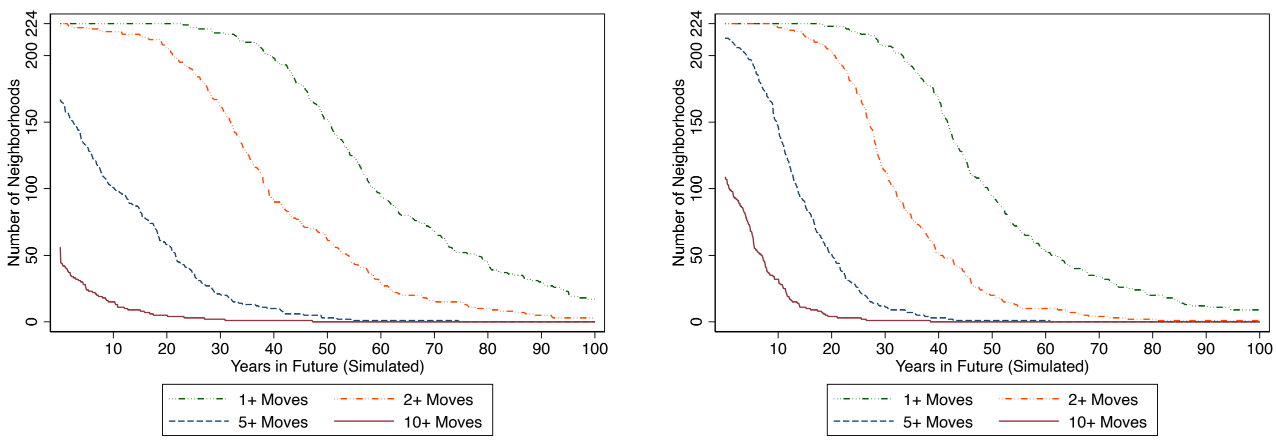


Notes: For each panel representing a different counterfactual, the arrows represent changes in simulated Dissimilarity Indices for households of each race and income from November 2004 to sorting equilibrium in the absence of exogenous shocks. Numbers in parentheses correspond to the relative change in dissimilarity. Gray arrows correspond to the baseline changes shown in Figure 3. A Black dissimilarity index of, say, 0.60, means that 60% of Black homeowners would have to be relocated in order to generate an equal distribution of Blacks across all Bay Area neighborhoods.

Figure 8: Number of Neighborhoods In Flux - Full Integration (Simulated)

(a) Baseline Responses and Moving Costs

(b) No Race or Income Discrimination



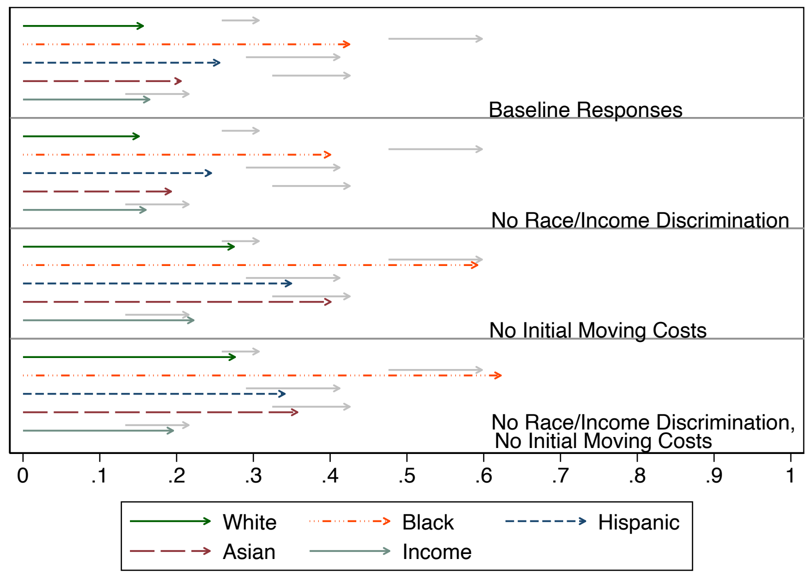
Notes: Each panel shows the number of neighborhoods with at least one, two, five or ten moves that change their socioeconomic composition (out of a total of 224 neighborhoods) under a different counterfactual. Simulation begins in November 2004.

5.3 The Role of the Initial Allocation of Households

We now consider a counterfactual that changes the initial allocation of households across neighborhoods; in particular, we re-allocate households so that all neighborhoods have the exact same initial socioeconomic compositions (mimicking a policy that generates full integration of all race and income groups). Because the compositions of neighborhoods as observed in November 2004 had likely arisen from an ongoing process of convergence to some equilibrium, this fully integrated counterfactual likely increases amenity mismatch. The first panel of Figure 8 plots the number of neighborhoods in flux after the full integration policy. As compared with the benchmark in Figure 2, this re-arrangement of households takes longer to reach sorting equilibrium. This is intuitive, as this policy likely leads to a major misalignment that takes longer to undo because of moving costs. Eliminating discrimination, as in the second panel of Figure 8, seems to speed up convergence slightly.

We explore the relationship between initial socioeconomic compositions and segregation dynamics in Figure 9 under four counterfactuals. When starting in a fully

Figure 9: Equilibrium Changes in Race and Income Segregation - Full Integration (Simulated)



Notes: For each panel representing a different counterfactual, the arrows represent changes in simulated Dissimilarity Indices for households of each race and income from November 2004 to sorting equilibrium in the absence of exogenous shocks. Numbers in parentheses correspond to the relative change in dissimilarity. Gray arrows correspond to the baseline changes shown in Figure 3. A Black dissimilarity index of, say, 0.60, means that 60% of Black homeowners would have to be relocated in order to generate an equal distribution of Blacks across all Bay Area neighborhoods.

integrated Bay Area, all dissimilarity indexes are zero by assumption. The top panel shows that this integration policy would reduce segregation in equilibrium (relative to the baseline counterfactual shown in the gray arrows). This is evidence of multiplicity of equilibria: initial conditions matter. With multiple equilibria, it is perhaps not surprising that segregation increases less under a full integration policy since moving costs prevent sorting by many households. The similarity of the effects in the top two panels implies that eliminating discrimination has very little impact on segregation. However, when moving costs are eliminated, segregation increases to levels that are closer to the levels of the gray arrows. This simply reflects the fact that without moving costs, the initial allocation of households is largely irrelevant since they can be reallocated costlessly. In the bottom panel, we eliminate both initial moving costs and discrimination. Households still reallocate costlessly, but they converge to an equilibrium that is a bit less segregated than in the third panel.

It is worth mentioning that the initial allocation of households is not entirely irrelevant even when moving costs are initially eliminated. The reason for this is that the friction from errors in households' expectations still exists (see, e.g., the first panel of Figure 6). Because the effects of discrimination are self-reinforcing, the timing of moves may impact the equilibrium allocation of households. This explains why the equilibrium under no initial moving costs starting from observed neighborhood compositions (the first panel of Figure 7) differs from the equilibrium under no initial moving costs starting from fully integrated neighborhoods (the third panel of Figure 9). This also explains why when we eliminate discrimination, the initial conditions become irrelevant (compare the second panel of Figure 7 with the fourth panel of Figure 9).

We conclude that moving costs are extremely important in explaining both the long-run level and speed of increases in segregation, as they restrict sorting by many households to their most desired neighborhoods. We would expect this in an environment where moving costs are large relative to discrimination as there is limited scope for changes in the race and income compositions of neighborhoods to trigger a cascade of new moves. This dampening of the feedback loop discussed in Schelling (1969) gives other amenities a more prominent role in explaining residential sorting and segregation.

6 House Prices and Segregation

6.1 Identification and Estimation

In this section, we explicitly incorporate neighborhood house prices into our framework to explore their impacts on segregation dynamics. We identify and estimate v_{gjt} and ϕ_g exactly as described in Section 3, so we restrict our discussion to the later stages after they are estimated. We augment our main regression equation (19) with a price term as follows:

$$v_{gjt} = \boldsymbol{\theta_g'} \boldsymbol{s^e_{jt}} + \alpha_g P_{jt} + \xi^P_{gjt} \tag{24}$$

where P_{jt} represents the average price of neighborhood j in period t, and α_g represents the causal effect of an increase in P_{jt} on the average cumulative indirect utility of group g households for neighborhood j in t. Note that $\xi^P_{gjt} = \xi_{gjt} - \alpha_g P_{jt}$, where ξ_{gjt} is the error from equation (10).

To close the model, it is now necessary to add a price equation

$$P_{jt} = \boldsymbol{\rho' s^e_{jt}} + \eta_{jt} \tag{25}$$

where η_{jt} is an error term. The coefficient $\boldsymbol{\rho}$ reflects the causal effect of a marginal change in $\boldsymbol{s^e_{jt}}$ on P_{jt}. Equation (25) merely captures the reduced-form, linear effect of s^e_{jt} on P_{jt}. For instance, it does not impose the assumptions required to interpret $\boldsymbol{\rho}$ as the marginal willingness-to-pay for $\boldsymbol{s^e_{jt}}$, as in the case of a standard static hedonic framework (Rosen (1974)).

Note that the parameters of equations (24) and (25) are related to the parameters of equation (10) as follows:

$$\beta_{g,g'} = \theta_{g,g'} + \rho_{g'} \cdot \alpha_g \tag{26}$$

where $\beta_{g,g'}$, $\theta_{g,g'}$ and $\rho_{g'}$ denote the g'-th element of the vectors $\boldsymbol{\beta_g}$, $\boldsymbol{\theta_g}$ and $\boldsymbol{\rho}$, respectively. Figure 10 describes their relationship graphically where each arrow represents a causal, direct effect that is captured by a parameter. The *total* causal effect of $s^e_{g'jt}$ on v_{gjt} ($\beta_{g,g'}$ from equation (10)) is simply equal to the direct effect holding prices constant ($\theta_{g,g'}$, from equation (24)) plus the indirect effect mediated through a change in price. This indirect effect arises through the interaction of two causal effects: the effect of $s^e_{g'jt}$ on P_{jt} ($\rho_{g'}$, from equation (25)) and the causal effect of P_{jt} on v_{gjt} (α_g, from equation (24)).

For intuition, it is useful to explicitly describe a hypothetical sequence of events that might be expected to occur within a single period t. Consider a change in $s^e_{g'jt}$. This may affect the desired inflows of households of different groups differently as reflected in $\theta_{g,g'}$. Next, P_{jt} may adjust to accommodate excess supply or demand in the neighborhood as reflected in $\rho_{g'}$, a reduced-form parameter that incorporates both supply-side and demand-side elasticities. As prices begin to adjust, the desired inflows of each group may respond differently according to α_g. This process may iterate a number of times so that by the end of t we observe P_{jt} and inflows_{gjt} for all g. For each g, the reduced-form coefficient $\beta_{g,g'}$ incorporates the effect of $\boldsymbol{s^e_{g'jt}}$ on v_{gjt} taking into account all of the adjustment processes that households might *expect* to occur during t, such as this

Figure 10: Different Channels of Causality from $\boldsymbol{s^{e}_{g'jt}}$ to v_{gjt}

$$\begin{tikzcd}
s^{e}_{g'jt} \arrow[r, "\theta_{g,g'}"] \arrow[d, "\rho_{g'}"] & v_{gjt} \\
P_{jt} \arrow[ur, "\alpha_g"] &
\end{tikzcd}$$

Notes: Each arrow in this graph represents a direct causal relationship. According to equation (26), the *total* causal effect of $s^{e}_{g'jt}$ on v_{gjt} ($\beta_{g,g'}$) is equal to the *direct* ($\theta_{g,g'}$) plus the *indirect* ($\rho_{g'} \cdot \alpha_g$) causal effects of $s^{e}_{g'jt}$ on v_{gjt}, where the indirect channel of causality is mediated via prices, P_{jt}.

one. We make no assumptions on the particulars of the process by which $s^{e}_{g'jt}$ causes v_{gjt}. We simply posit that whatever households expect the process to be, it can be decomposed into one price channel, $\rho_{g'} \cdot \alpha_g$ and a second residual channel $\theta_{g,g'}$.[31]

In our main analysis we simply identified $\boldsymbol{\beta_g}$. This allowed us to characterize segregation without imposing additional assumptions required to disentangle the direct and indirect effects of $\boldsymbol{s_{jt}}$ on v_{gjt}. We intentionally selected the set of endogenous amenities parsimoniously by focusing on the two *primitive* endogenous dimensions along which households sort: race and income. However, that specification was not equipped to study the potential role of prices in explaining segregation. To do so, we must separately identify $\theta_{g,g'}$, α_g and $\rho_{g'}$ for all g and g', which requires an additional assumption.

Assumption 4. *For some $T > T' \geq 1$,*

1. $COV\left(\xi^{P}_{gjt}, \boldsymbol{s_{jt-T}} | \mathit{inflows}_{\boldsymbol{jt-1}}, ..., \mathit{inflows}_{\boldsymbol{jt-T'}}\right) = 0.$

2. $COV\left(\xi^{P}_{gjt}, P_{jt-T} | \mathit{inflows}_{\boldsymbol{jt-1}}, ..., \mathit{inflows}_{\boldsymbol{jt-T'}}\right) = 0.$

Following Assumption 4, we use $\boldsymbol{s_{jt-T}}$ and P_{jt-T} as IVs for $\boldsymbol{s_{jt}}$ and P_{jt} to estimate $\boldsymbol{\theta_g}$ and α_g in the Two Stage Least Squares (2SLS) regression

[31] Note, in particular, that there is no need to assume that supply equals demand within a given period.

$$v_{gjt} = \boldsymbol{\theta_g'} \boldsymbol{s_{jt}} + \alpha_g P_{jt} + \boldsymbol{\Lambda_g} \left(\boldsymbol{\hat{v}_{jt-1}, ..., \hat{v}_{jt-T'}} \right) + \gamma_{gt} + \tilde{\xi}^P_{gjt}, \tag{27}$$

where $\tilde{\xi}^P_{gjt} = \left(\xi^P_{gjt} - \boldsymbol{\Lambda_g} \left(\boldsymbol{\hat{v}_{jt-1}}, \ldots, \boldsymbol{\hat{v}_{jt-T'}} \right) \right) + \boldsymbol{\beta}_g' \left(\boldsymbol{s^e_{jt}} - \boldsymbol{s_{jt}} \right) + (\hat{v}_{gjt} - v_{gjt})$.[32]

Assumption 4.1 is a clear analog to Assumption 3. It simply replaces ξ_{gjt} with ξ^P_{gjt}, so it further assumes that a subcomponent of the unobservable ξ_{gjt}, namely ξ^P_{gjt}, is also uncorrelated to the IV $\boldsymbol{s_{jt-T}}$. Assumption 4.2 is also similar in spirit to Assumption 3, but there are key differences. It states that no unobservable affecting inflow decisions in t should be correlated to P_{jt-T} once we condition on $\text{inflows}_{\boldsymbol{jt-1}}, ..., \text{inflows}_{\boldsymbol{jt-T'}}$. It is useful to understand which *additional* source of variation (independent of $\boldsymbol{s_{jt-T}}$) the IV P_{jt-T} exploits. Consider two neighborhoods j and j' that are otherwise identical in $t-T$, except for the fact that j has a higher level of one amenity that is liked (in different intensities) by households of all groups. These different intensities may lead to $\boldsymbol{s_{jt-T}} \neq \boldsymbol{s_{j't-T}}$, which contributes to the relevance of the IV $\boldsymbol{s_{jt-T}}$ as discussed in Section 3. However, the common component to the valuation of this amenity across all groups would not lead to $\boldsymbol{s_{jt-T}} \neq \boldsymbol{s_{j't-T}}$ yet would lead to $P_{jt-T} > P_{j't-T}$ because of an excess demand for neighborhood j relative to neighborhood j' since all groups tend to like this amenity.[33] Only those amenities that do not affect inflow decisions in $t-1, ..., t-T'$ are exploited for identification since we control for $\text{inflows}_{\boldsymbol{jt-1}}, ..., \text{inflows}_{\boldsymbol{jt-T'}}$.

Why would P_{jt-T} be correlated with P_{jt} conditional on $\text{inflows}_{\boldsymbol{jt-1}}, ..., \text{inflows}_{\boldsymbol{jt-T'}}$? In Section 3.2, we discussed that the relevance of $\boldsymbol{s_{jt-T}}$ as an IV for $\boldsymbol{s_{jt}}$ arises from an asymmetry on the demand-side between $\boldsymbol{s_{jt}}$, a stock variable that depends on decisions from the distant past, and v_{gjt}, a flow variable that depends only on inflow decisions in period t. The source of relevance of P_{jt-T} as an IV for P_{jt} arises from another asymmetry: while the outcome variable v_{gjt} depends on unobservables only via ξ^P_{gjt}, the endogenous variable P_{jt} depends on unobservables ξ^P_{gkt} for $k \neq j$. This is because all neighborhoods compete with each other in the housing market, so amenities ξ^P_{gkt} in neighborhoods $k \neq j$ may impact the prices of neighborhood j (Berry, Levinsohn and Pakes (1995)). To the extent that these amenities may persist over time, P_{jt-T} might be correlated to P_{jt}. The idea underlying Assumption 3 is that by controlling for inflows from $t-1$ to $t-T'$, the component of ξ^P_{gkt-T} that is correlated to ξ^P_{gjt-T} will plausibly not persist to t. Still, some of ξ^P_{gkt-T} may persist to t without affecting inflows

[32] See the discussion around equation (20) for an explanation of why $\boldsymbol{s_{jt-T}}$ is uncorrelated to $\tilde{\xi}^P_{gjt} - \xi^P_{gjt}$. An analogous argument implies P_{jt-T} is also uncorrelated to $\tilde{\xi}^P_{gjt} - \xi^P_{gjt}$.

[33] More formally, let $A_{jt-\tau}$ be the amenity in this example for $\tau > T$. Then $A_{jt-\tau} - \mathbb{E}\left[A_{jt-\tau} | \boldsymbol{s_{jt-T}}\right]$ is the common component of this amenity that would not affect $\boldsymbol{s_{jt-T}}$ yet would affect P_{jt-T}.

from $t-1$ to $t-T'$ but rather by affecting P_{jt} via competition across neighborhoods.

With estimates of $\boldsymbol{\beta}$, $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$, obtained via the validity Assumptions 3 and 4, we can obtain estimates of $\boldsymbol{\rho}$. To see this, note that Assumptions 3 and 4.1 imply $COV\left(\eta_{jt}, \boldsymbol{s_{jt-T}} | \text{inflows}_{\boldsymbol{jt-1}}, ..., \text{inflows}_{\boldsymbol{jt-T'}}\right) = 0$, where η_{jt} is the error from equation (25).[34] This allows us to estimate $\boldsymbol{\rho}$ via 2SLS using $\boldsymbol{s_{jt-T}}$ as an IV for $\boldsymbol{s_{jt}}$ in the equation

$$P_{jt} = \boldsymbol{\rho' s_{jt}} + \boldsymbol{\Gamma}(\boldsymbol{v_{jt-1}}, ..., \boldsymbol{v_{jt-T'}}) + \lambda_t + \tilde{\eta}_{jt} \tag{28}$$

where $\tilde{\eta}_{gjt} = (\eta_{gjt} - \boldsymbol{\Gamma}\left(\hat{\boldsymbol{v}}_{\boldsymbol{jt-1}}, \ldots, \hat{\boldsymbol{v}}_{\boldsymbol{jt-T'}}\right)) + \boldsymbol{\rho}'\left(\boldsymbol{s^{e}_{jt}} - \boldsymbol{s_{jt}}\right)$ and $\boldsymbol{\Gamma}(\cdot)$ is a flexible function, specified analogously to $\boldsymbol{\Lambda_g}(\cdot)$.

With estimates of $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\phi}}$ and $\hat{\boldsymbol{\rho}}$, we augment the simulation procedure in a straightforward manner. As before, we simulate the trajectory $\mathbb{T}_t(\tilde{\boldsymbol{s}}) = \{\mathbb{T}^0_t(\tilde{s}), \mathbb{T}^1_t(\tilde{s}), ...\}$ where $\mathbb{T}^0_t(\tilde{s}) = \tilde{\boldsymbol{s}}$, and $\mathbb{T}^{\tau}_t(\tilde{\boldsymbol{s}}) = \boldsymbol{s}_t\left(\mathbb{T}^{\tau-1}_t(\tilde{\boldsymbol{s}})\right)$. However, now the function $\boldsymbol{s}_t(\cdot)$ changes slightly relative to what we had in Section 3. We replace equation (21) with

$$v_{gjt}\left(\tilde{\boldsymbol{s}}\right) = \hat{v}_{gjt} + \hat{\boldsymbol{\theta}}'_{\boldsymbol{g}}\left(\tilde{\boldsymbol{s}}_{\boldsymbol{j}} - \boldsymbol{s_{jt}}\right) + \hat{\alpha}_g\left(P_{jt}\left(\tilde{\boldsymbol{s}}\right) - P_{jt}\right) \tag{29}$$

and we add one more equation

$$P_{jt}\left(\tilde{\boldsymbol{s}}\right) = P_{jt} + \hat{\boldsymbol{\rho}}'\left(\tilde{\boldsymbol{s}}_{\boldsymbol{j}} - \boldsymbol{s_{jt}}\right) \tag{30}$$

Thus, instead of simulating the trajectory using equations (21), (22) and (23), we simulate it using equations (29), (30), (22) and (23).

Remark 4. As with the interpretation of $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ incorporates two components: $\alpha_g = \frac{\partial v_{gjt}}{\partial P_{jt}} = \frac{\partial u_{gjt}}{\partial P_{jt}} + \frac{\partial CV_{gjt}}{\partial P_{jt}}$. The flow utility component $(\frac{\partial u_{gjt}}{\partial P_{jt}})$ should be negative since households prefer to pay lower prices for their house, all else constant. However, the sign of the continuation value component $(\frac{\partial CV_{gjt}}{\partial P_{jt}})$ is theoretically ambiguous since it depends on whether the price of a neighborhood today signals disproportionate expected future appreciation than an otherwise comparable neighborhood, which would have consequences for homeowners' expected wealth.[35]

[34] By substituting equation (25) into equation (24), we obtain $v_{gjt} = \boldsymbol{\theta'_g s^{e}_{jt}} + \boldsymbol{\alpha_g \rho' s^{e}_{jt}} + \eta_{jt}\alpha_g + \xi^{P}_{gjt}$. Comparing this to equation (10), we conclude that $\xi_{gjt} = \eta_{jt}\alpha_g + \xi^{P}_{gjt}$. It follows that Assumptions 3 and 4.1 imply $COV\left(\eta_{jt}, \boldsymbol{s_{jt-T}} | \text{inflows}_{\boldsymbol{jt-1}}, ..., \text{inflows}_{\boldsymbol{jt-T'}}\right) = 0$.

[35] The buying and selling of a house may impact household wealth. Despite its undeniable importance when studying the behavior of homeowners, we do not explicitly model the effects of moving on wealth, and we do not allow for household heterogeneity by wealth, as done by Bayer et al. (2016) for White homeowners. In our context, doing so would substantially increase the number of groups of households that we would need to consider and would render our analysis infeasible since there are not enough

6.2 Empirics

We compute monthly neighborhood prices by averaging the sales prices of all transactions that we observe. The average neighborhood price in our sample is \$329,000 with a standard deviation of \$232,000. There is considerable appreciation over our sample period, as the average price rises from \$248,000 in 1990 to \$564,000 in 2004 (all prices in constant November 2004 dollars). For practical purposes, we impose one additional restriction on $\boldsymbol{\alpha}$, namely that these parameters vary only by income, not by race and income (e.g., rich Whites and rich Blacks have the same $\boldsymbol{\alpha_g}$.) We do so only because otherwise we cannot obtain precise estimates of $\boldsymbol{\alpha}$.

We present estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$ in Table 3. Compared to Table 2, we conclude that our estimates of responses to socioeconomic compositions with and without prices ($\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ respectively) are quite similar, which suggests a limited role for prices in influencing segregation dynamics. Households of both income groups respond negatively to higher neighborhood prices, but the poor are over four times more price-sensitive than the rich.

We present estimates of $\boldsymbol{\rho}$ in Table 4. They imply that a 10 percentage point increase in the expected Black share of a neighborhood, all else constant, leads to a reduction in average price of \$19,400. This effect is over twice as large for the same increase in the expected Hispanic or poor shares of a neighborhood and roughly the same size for the same increase in the expected Asian share of a neighborhood. We should not interpret these estimates as households' marginal willingness to pay for their neighbors as discussed in Bishop and Murphy (2019). Rather, we simply use these estimates to simulate how households expect neighborhood prices to change depending on endogenous changes in the expected socioeconomic composition.[36]

In Figure 11, we present simulated trajectories of segregation levels when price is explicitly incorporated into the analysis. Comparing the main arrows in the top panel

households of each race and income level to study their decisions by wealth levels. Note, however, that wealth is incorporated in our analysis, even if implicitly, since the moving cost parameters ϕ_g are allowed to vary by groups, and these groups may have different wealth on average.

[36]All IV estimates of $\boldsymbol{\theta}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\rho}$ are robust to our choice of T analogous to what we show in Figures 13 and 14 for $\boldsymbol{\beta}$. Larger values of T imply weaker Assumptions 3 and 4, since, all else constant, amenities from $t-T$ or before would be less likely to affect inflows in t. Hence, the stability of our estimates is evidence in favor of our exclusion restrictions. These results are available upon request.

Table 3: Responses to the Socioeconomic Compositions ($\boldsymbol{\theta}$) and Prices ($\boldsymbol{\alpha}$) of Neighborhoods

	White		Black		Hispanic		Asian	
	Rich	Poor	Rich	Poor	Rich	Poor	Rich	Poor
Responses to:								
Black Share	-2.95***	-2.64***	3.36***	2.08***	0.35	-0.54	-1.29***	-1.16**
	(0.40)	(0.45)	(0.43)	(0.42)	(0.40)	(0.43)	(0.44)	(0.41)
Hispanic Share	-4.75***	-2.09	-1.63**	-4.17***	11.23***	8.01***	-0.46	-0.92
	(0.67)	(0.92)	(0.67)	(0.69)	(0.82)	(0.82)	(0.78)	(0.80)
Asian Share	-0.52	-4.21***	0.26	-1.44***	-0.77	-2.18***	4.98***	7.09***
	(0.48)	(0.63)	(0.52)	(0.48)	(0.61)	(0.62)	(0.67)	(0.69)
Poor Share	-0.86***	2.80***	-2.52***	-0.37	-1.84***	-0.63*	-3.12***	0.61***
	(0.26)	(0.44)	(0.38)	(0.30)	(0.39)	(0.36)	(0.38)	(0.41)
Response to Price† (Millions)	-0.39**	-1.71***	-0.39**	-1.71***	-0.39**	-1.71***	-0.39**	-1.71***
	(0.19)	(0.20)	(0.19)	(0.20)	(0.19)	(0.20)	(0.19)	(0.20)
R^2	0.79							
Num. of Observations	147,840							

Notes: This table shows 2SLS estimates of $\boldsymbol{\theta}_g$ and α_g from equation (27). The specification includes group-month fixed effects and control variables $\boldsymbol{\Lambda_g}\,(\boldsymbol{\hat{v}_{jt-1}}, ..., \boldsymbol{\hat{v}_{jt-12}})$ (see Footnote 25). As instrumental variables, we use $s_{g'jt-13}$ for all g' as well as P_{jt-13}. White is the omitted racial share and rich is the omitted income share. All standard errors clustered by group-month. * - 90% significance, ** - 95% significance, *** - 99% significance. The p-values for both the Cragg-Donald and the Kleinbergen-Paap weak identification tests are less than 0.01, which implies a strong first stage. † α_g is allowed to vary only by income groups in this specification.

Table 4: Implicit Price of $\boldsymbol{s^e_{jt}}$ ($\boldsymbol{\rho}$)

	Black Share	Hispanic Share	Asian Share	Poor Share
Price (Millions)	-0.201***	-0.565***	-0.165***	-0.472***
	(0.010)	(0.018)	(0.020)	(0.017)
R^2		0.62		
Num. Obs.		36,960		

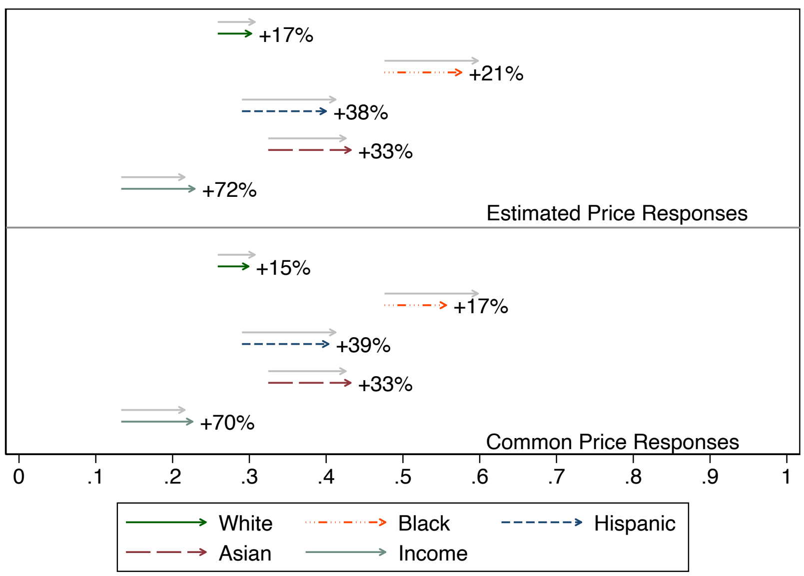
Notes: This table shows 2SLS estimates of $\boldsymbol{\rho}$ from equation (28). The specification includes month fixed effects and control variables $\boldsymbol{\Gamma}(\boldsymbol{\hat{v}_{jt-1}}, ..., \boldsymbol{\hat{v}_{jt-12}})$ (see Footnote 25). As instrumental variables, we use $s_{g'jt-13}$ for all g'. White is the omitted racial share and rich is the omitted income share. All standard errors clustered by month. * - 90% significance, ** - 95% significance, *** - 99% significance. The p-values for both the Cragg-Donald and the Kleinbergen-Paap weak identification tests are less than 0.01, which implies a strong first stage.

with the gray arrows representing our baseline findings, we conclude that the results are essentially unchanged. Indeed, the results of all counterfactual analyses carried out in Section 5 are almost identical when prices are explicitly accounted for. If we were conducting these simulations with the true values of $\boldsymbol{\beta}$, $\boldsymbol{\theta}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\rho}$ (i.e., not estimated) then this would be expected because of the identity expressed in equation (26). Because we estimate these causal parameters, our findings could in principle have been different under violations of Assumptions 3, 4, or even our restriction that α_g can only vary across income levels. In light of this, we view the similarity of our findings as reassuring evidence in support of our identification strategy.

The explicit inclusion of prices provides the opportunity to consider a new counterfactual whereby all households have an identical response to prices (i.e., α_g is the same for all g). Specifically, we set α for each group to be equal to the population weighted average of $\hat{\alpha}_g$ across all g. A comparison of the bottom panel and the top panel of Figure 11 allows us to infer the role of heterogeneous price responses on segregation. It is clear that this heterogeneity has a very small effect on segregation on the same order of magnitude as race and income discrimination.

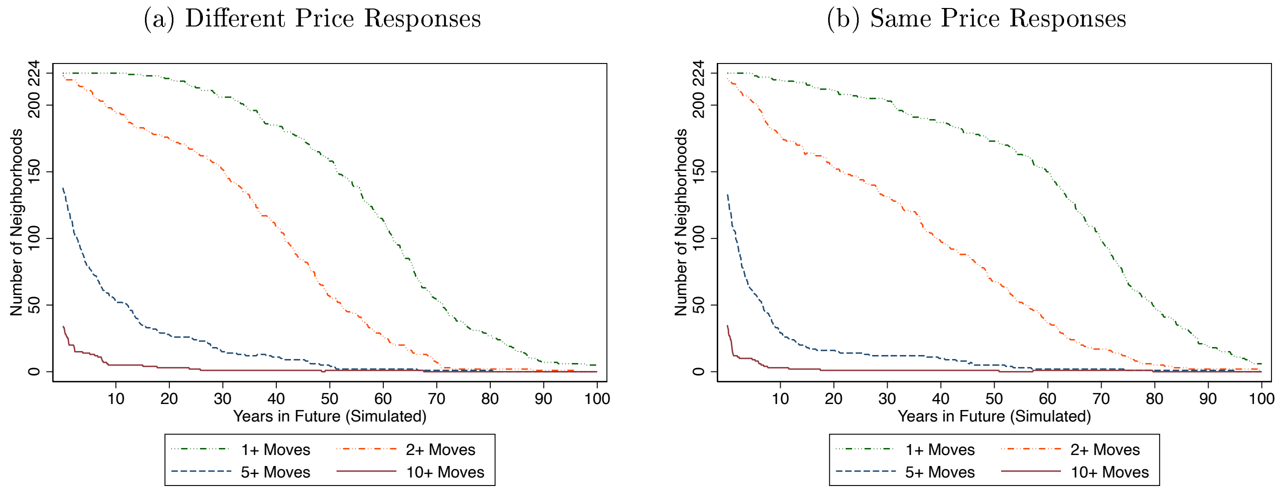
In Figure 12, we compare the turnover of neighborhoods in the baseline case (first panel) versus in the counterfactual case where everyone has the same price response (second panel). As expected, the first panel of Figure 12 is very similar to Figure 2, which only implicitly incorporates prices into the analysis. Moreover, the first and second panels of Figure 12 look very similar, suggesting that heterogeneous price responses have little impact on neighborhood dynamics as well.

Figure 11: Trajectories of Segregation Levels by Race and by Income - Explicit Price Responses



Notes: For each panel representing a different counterfactual, the arrows represent changes in simulated Dissimilarity Indices for households of each race and income from November 2004 to sorting equilibrium in the absence of exogenous shocks. Numbers in parentheses correspond to the relative change in dissimilarity. Gray arrows correspond to the baseline changes shown in Figure 3. A Black dissimilarity index of, say, 0.60, means that 60% of Black homeowners would have to be relocated in order to generate an equal distribution of Blacks across all Bay Area neighborhoods. The rich dissimilarity index is identical to the poor dissimilarity index.

Figure 12: Number of Neighborhoods In Flux (Simulated)



Notes: Figure shows the number of neighborhoods with at least one, two, five or ten moves that change their socioeconomic composition (out of a total of 224 neighborhoods). The second panel is under the counterfactual where α_g is equal to the population weighted average of $\hat{\boldsymbol{\alpha}}$. Simulation begins in November 2004.

7 Conclusion

Neighborhoods constantly evolve: their amenities are not static and their residents are in flux. Disequilibrium models of segregation tend to attribute this evolution to endogenous changes in neighborhood residents arising from discrimination, while disaggregated models of residential choice tend to attribute this evolution to exogenous changes in other amenities arising from serially correlated external shocks. In this paper, we develop an empirical framework that bridges these two approaches and provides new perspectives on how the aggregate phenomenon of segregation arises from the accumulation of disaggregate residential choices.

We use this framework to study the determinants of race and income segregation in the San Francisco Bay Area from 1990 to 2004. By delineating the interconnected roles of socioeconomic discrimination, other neighborhood amenities, incomplete information, moving costs, initial allocations of households across neighborhoods, and heterogeneity in price-sensitivity, we explore the underlying forces that drive segregation through counterfactual analyses. We find that while discrimination and heterogeneous price responses matter for segregation, they are much less important than sorting on the basis of other amenities. This is in large part due to frictions, primarily moving costs (although incomplete information also plays a discernible role). These frictions prevent much desired sorting from occurring, which reduces the scope for the feedback loop mechanism generated by endogenous discriminatory sorting. The interplay of all of these forces contribute to a metropolitan area that is on the path to further segregation though much less than in the absence of frictions.

An important caveat in our analysis is that we do not observe the socioeconomic composition of renters over time. This may be less damaging to our conclusions if the aspects of the expected composition of neighborhoods that are most relevant to sorting decisions are the ones proxied by the actual composition of homeowners (e.g., different allocations of local public goods spending depending on the socioeconomic composition of local taxpayers). However, this may be a concern in neighborhoods with lower rates of homeownership if the aspects of the expected composition of neighborhoods that are most relevant to sorting decisions are the compositions of the people that *use* public goods and, at the same time, landlords' socioeconomic status is a poor predictor of tenants' socioeconomic status. In any case, because renters face relatively lower moving costs than homeowners, we would expect to find patterns of segregation somewhere in between our baseline findings and our counterfactual findings with zero moving costs. Future research with access to better data is needed to address these issues.

Richer data would also provide opportunities to study the roles of additional frictions in shaping segregation. For instance, supply-driven discrimination might disproportionately prevent racial minorities and the poor from sorting to their desired neighborhoods (e.g., Timmins and Christensen (2018)). It would also allow us to explore socioeconomic segregation at finer levels; for instance, we could consider more income groups or we could disaggregate Asians into Chinese-Americans, Korean-Americans, etc. Ultimately, we view our framework as a platform for the empirical analysis of determinants of segregation that can be directly adapted to various contexts. The use of this framework to study sorting along different demographic dimensions (e.g., race, income, partisanship, education) in different settings (e.g., neighborhoods, schools, virtual communities, physical venues) could prove valuable in revealing the importance of different cleavages in our society.

References

Ackerberg, Daniel A. 2020. "Timing Assumptions and Effi ciency: Empirical Evidence in a Production Function Context.".

Aguirregabiria, Victor and Pedro Mira. 2010. "Dynamic discrete choice structural models: A survey." *Journal of Econometrics* 156(1):38–67.

Arellano, Manuel and Stephen Bond. 1991. "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations." *The review of economic studies* 58(2):277–297.

Banzhaf, H. Spencer and Randall P. Walsh. 2013. "Segregation and Tiebout sorting: The link between place-based investments and neighborhood tipping." *Journal of Urban Economics* 74(0):83 – 98.

Bayer, P. and C. Timmins. 2005. "On the equilibrium properties of locational sorting models." *Journal of Urban Economics* 57(3):462–477.

Bayer, P. and C. Timmins. 2007. "Estimating Equilibrium Models Of Sorting Across Locations." *The Economic Journal* 117(518):353–374.

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.

Bayer, P., N. Keohane and C. Timmins. 2009. "Migration and hedonic valuation: The case of air quality." *Journal of Environmental Economics and Management* 58(1):1–14.

Bayer, P., R. McMillan and K. Rueben. 2004*a*. An equilibrium model of sorting in an urban housing market. Technical report National Bureau of Economic Research.

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

Bayer, Patrick, Robert McMillan and Kim S. Rueben. 2004*b*. "What drives racial segregation? New evidence using Census microdata." *Journal of Urban Economics*

56(3):514 – 535.
URL: *http://www.sciencedirect.com/science/article/pii/S0094119004000634*

Becker, G.S. and K.M. Murphy. 2000. *Social Economics: Market Behavior in a Social Environment*. Harvard University Press.

Bergman, Peter. 2016. "The Effects of School Integration: Evidence from a Randomized Desegregation Program." *Working Paper* .

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

Berry, S.T. 1994. "Estimating Discrete-Choice Models of Product Differentiation." *The RAND Journal of Economics* pp. 242–262.

Bishop, Kelly C and Alvin D Murphy. 2019. "Valuing time-varying attributes using the hedonic model: when is a dynamic approach necessary?" *Review of Economics and Statistics* 101(1):134–145.

Caetano, Gregorio. 2019. "Neighborhood sorting and the value of public school quality." *Journal of Urban Economics* 114:103193.

Caetano, Gregorio and Vikram Maheshri. 2017. "School segregation and the identification of tipping behavior." *Journal of Public Economics* 148:115–135.

Caetano, Gregorio and Vikram Maheshri. 2020. Explaining Recent Trends in US School Segregation. Technical report Working Paper.

Card, D., A. Mas and J. Rothstein. 2008a. "Tipping and the Dynamics of Segregation." *Quarterly Journal of Economics* 123(1):177–218.

Card, David and Jesse Rothstein. 2007. "Racial segregation and the black–white test score gap." *Journal of Public Economics* 91(11):2158–2184.

Chetty, Raj, Nathaniel Hendren, Patrick Kline and Emmanuel Saez. 2014. Where is the land of opportunity? The geography of intergenerational mobility in the United States. Technical report National Bureau of Economic Research.

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

Epple, D. and H. Sieg. 1999. "Estimating Equilibrium Models of Local Jurisdictions." *Journal of Political Economy* 107(4).

Epple, D., T. Romer and H. Sieg. 2001. "Interjurisdictional sorting and majority rule: An empirical analysis." *Econometrica* 69(6):1437–1465.

Epple, D., T. Romer and H. Sieg. 2003. "Interjurisdictional sorting and majority rule: An empirical analysis." *Econometrica* 69(6):1437–1465.

Epple, Dennis, Radu Filimon and Thomas Romer. 1984. "Equilibrium among local jurisdictions: toward an integrated treatment of voting and residential choice." *Journal of Public Economics* 24(3):281–308.

Ferreira, Fernando. 2010. "You can take it with you: Proposition 13 tax benefits, residential mobility, and willingness to pay for housing amenities." *Journal of Public Economics* 94(9):661–673.
URL: *https://www.sciencedirect.com/science/article/pii/S0047272710000435*

Guryan, Jonathan. 2004. "Desegregation and black dropout rates." *The American Economic Review* 94(4):919–943.

Kiel, Katherine A and Jeffrey E Zabel. 1996. "House price differentials in US cities: Household and neighborhood racial effects." *Journal of housing economics* 5(2):143–165.

Kling, Jeffrey R, Jens Ludwig and Lawrence F Katz. 2005. "Neighborhood effects on crime for female and male youth: Evidence from a randomized housing voucher experiment." *The Quarterly Journal of Economics* pp. 87–130.

Kuminoff, Nicolai V, V Kerry Smith and Christopher Timmins. 2013. "The new economics of equilibrium sorting and policy evaluation using housing markets." *Journal of Economic Literature* 51(4):1007–1062.

Ludwig, Jens, Greg J Duncan, Lisa A Gennetian, Lawrence F Katz, Ronald C Kessler, Jeffrey R Kling and Lisa Sanbonmatsu. 2012. "Neighborhood effects on the long-term well-being of low-income adults." *Science* 337(6101):1505–1510.

Manski, Charles F. 2004. "Measuring expectations." *Econometrica* 72(5):1329–1376.

Mason, Susan M, Lynne C Messer, Barbara A Laraia and Pauline Mendola. 2009. "Segregation and preterm birth: the effects of neighborhood racial composition in North Carolina." *Health & Place* 15(1):1–9.

Mouw, Ted and Barbara Entwisle. 2006. "Residential Segregation and Interracial Friendship in Schools1." *American Journal of sociology* 112(2):394–441.

Olley, Steven and Ariel Pakes. 1996. "The dynamics of productivity in the telecomunications equipment industry." *Econometrica* 64:1263–97.

Ringo, Daniel. 2013. "Residential Clustering by Race and Income: What Drives the Sorting?" *Mimeo* .

Rivkin, Steven G, Eric A Hanushek and John F Kain. 2005. "Teachers, schools, and academic achievement." *Econometrica* 73(2):417–458.

Rosen, Sherwin. 1974. "Hedonic prices and implicit markets: product differentiation in pure competition." *Journal of political economy* 82(1):34–55.

Schelling, Thomas C. 1969. "Models of Segregation." *The American Economic Review* 59(2):488–493.

Schelling, Thomas C. 1971. "Dynamic Models of Segregation." *Journal of Mathematical Sociology* 1:143–186.

Tiebout, Charles. 1956. "A Pure Theory of Local Expenditures." *The Journal of Political Economy* 64(5):416–424.

Timmins, Christopher and Peter Christensen. 2018. "Sorting or Steering: Experimental Evidence on the Economic Effects of Housing Discrimination." *Working Paper* .

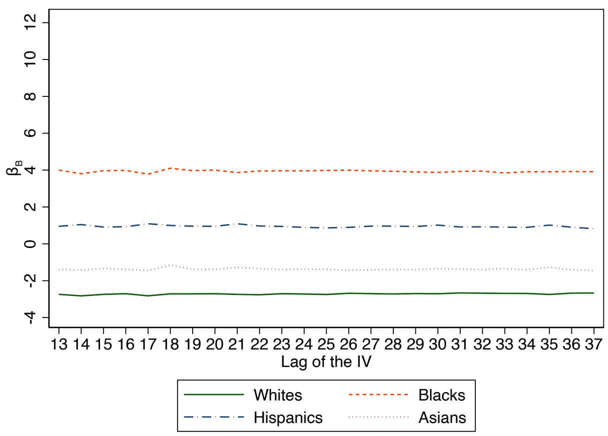
A Appendix Tables and Figures

Table 5: OLS Estimates of Responses to the Race and Income Compositions of Neighborhoods ($\boldsymbol{\beta}$)

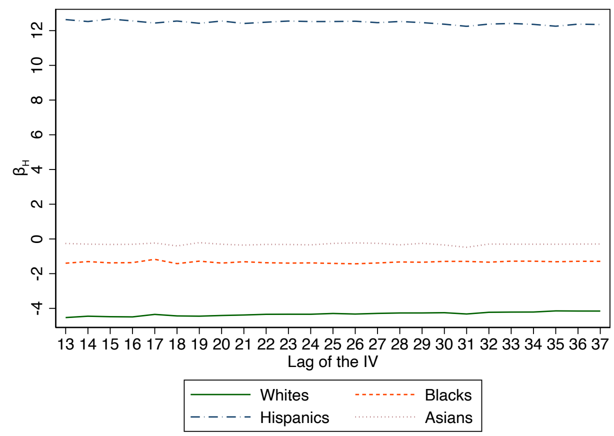
	White		Black		Hispanic		Asian	
	Rich	Poor	Rich	Poor	Rich	Poor	Rich	Poor
Responses to:								
Black Share	-12.25***	-7.32***	13.47***	16.35***	-0.81***	8.49***	-4.81***	1.02***
	(0.35)	(0.52)	(0.43)	(0.43)	(0.33)	(0.35)	(0.37)	(0.34)
Hispanic Share	-17.69***	2.66***	8.18***	8.86***	35.92***	43.64***	-1.76***	16.25***
	(0.51)	(0.53)	(0.47)	(0.38)	(0.67)	(0.53)	(0.46)	(0.48)
Asian Share	-5.63***	-9.99***	-0.44	-2.64***	-0.21	-3.54***	24.24***	26.63***
	(0.30)	(0.38)	(0.34)	(0.28)	(0.35)	(0.37)	(0.56)	(0.61)
Poor Share	-7.01***	1.28***	-7.52***	1.11***	-12.77***	-2.60***	-16.17***	-2.91***
	(0.35)	(0.39)	(0.30)	(0.27)	(0.30)	(0.35)	(0.37)	(0.34)
R^2	0.38							
Num. of Observations	147,840							

Notes: This specification includes only group-month fixed effects as controls. White is the omitted racial share and rich is the omitted income share. All standard errors clustered by group-month. * - 90% significance, ** - 95% significance, *** - 99% significance.

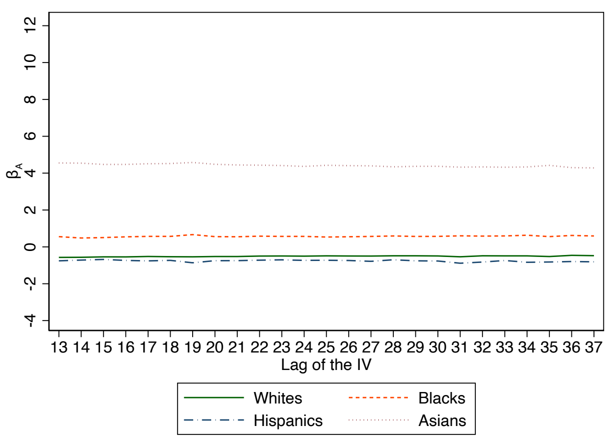
Figure 13: Responses of Rich Households of Different Races to Race and Income Compositions for Different Values of T



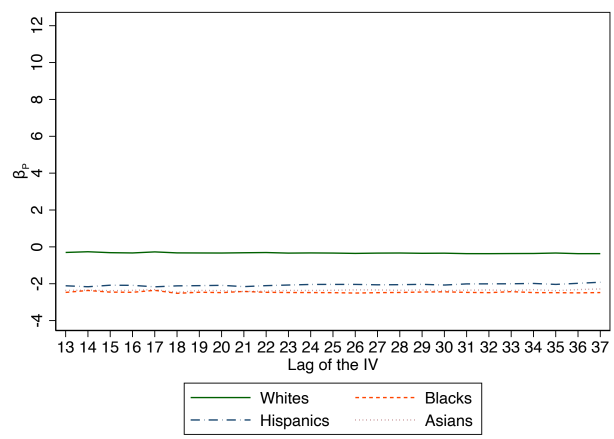
(a) Responses to Black Share



(b) Responses to Hispanic Share



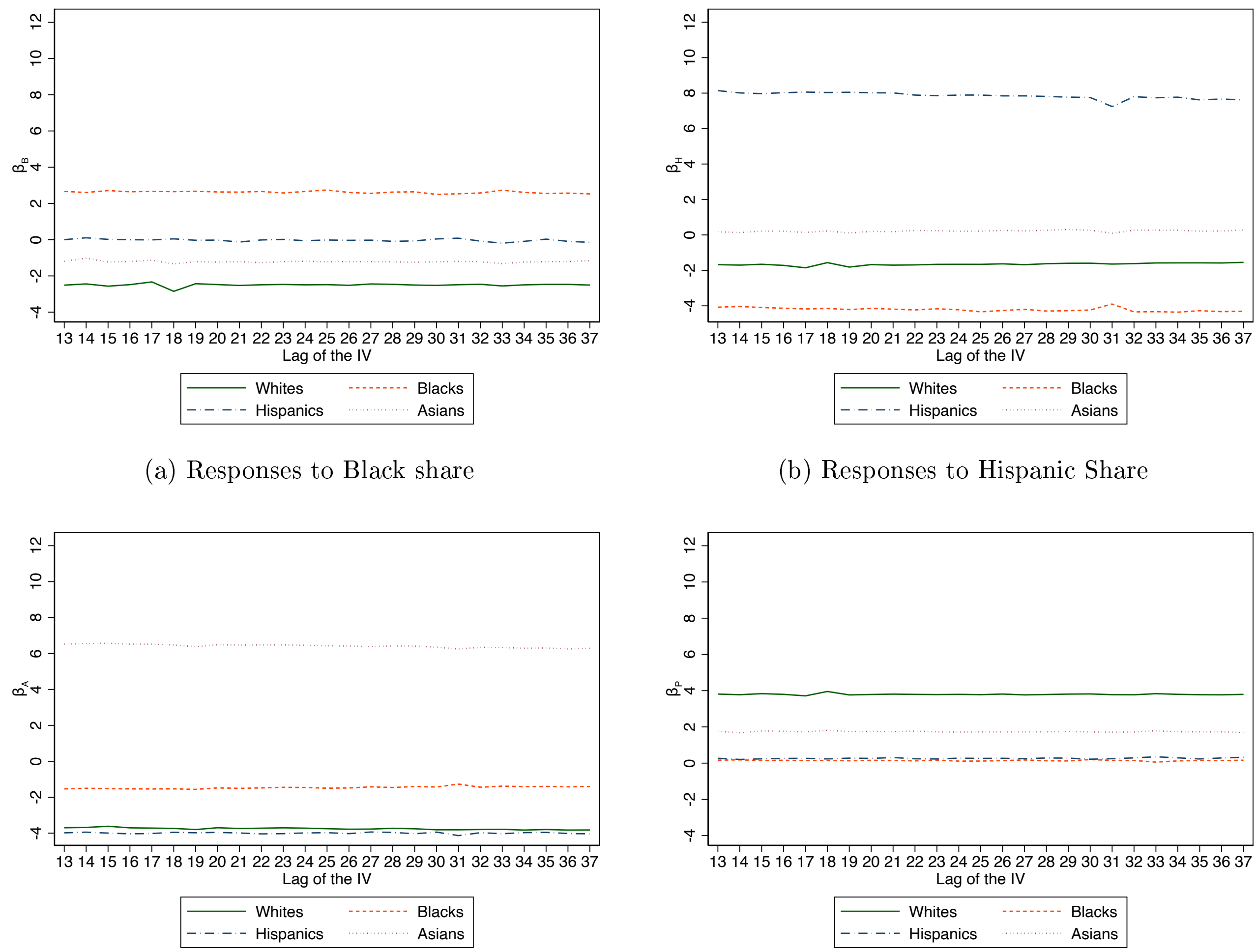
(c) Responses to Asian Share



(d) Responses to Poor Share

Notes: Each panel shows $\hat{\beta}_{g,g'}$ for all g' for different values of T, the lag of the Instrumental Variable, $\boldsymbol{s_{jt-T}}$. In all panels of this figure g represents rich Whites, Blacks, Hispanics and Asians. We set $T' = 12$ in all cases. For comparison, we keep the sample constant in all cases, so the first 37 months of the sample are not used in the estimation of $\boldsymbol{\beta}$ for all values of T.

Figure 14: Responses of Poor Households of Different Races to Race and Income Compositions for Different Values of T



(a) Responses to Black share

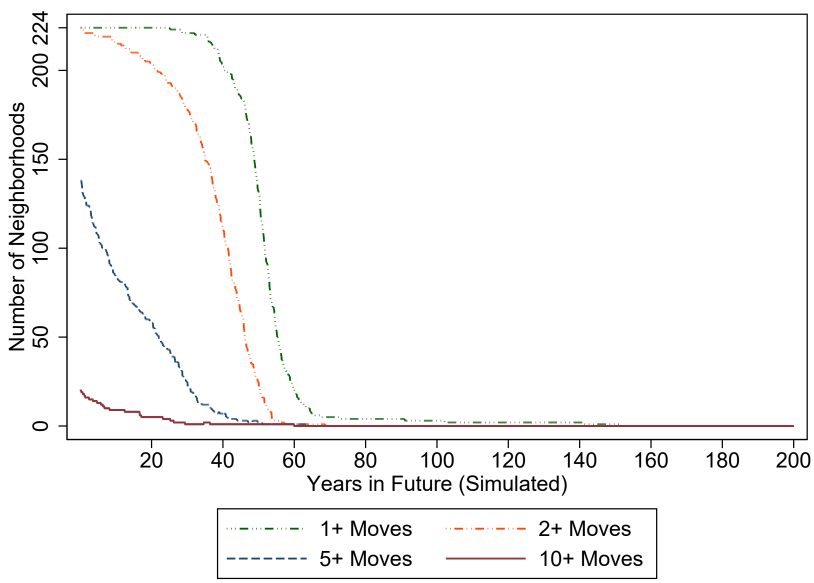
(b) Responses to Hispanic Share

(c) Responses to Asian Share

(d) Responses to Poor Share

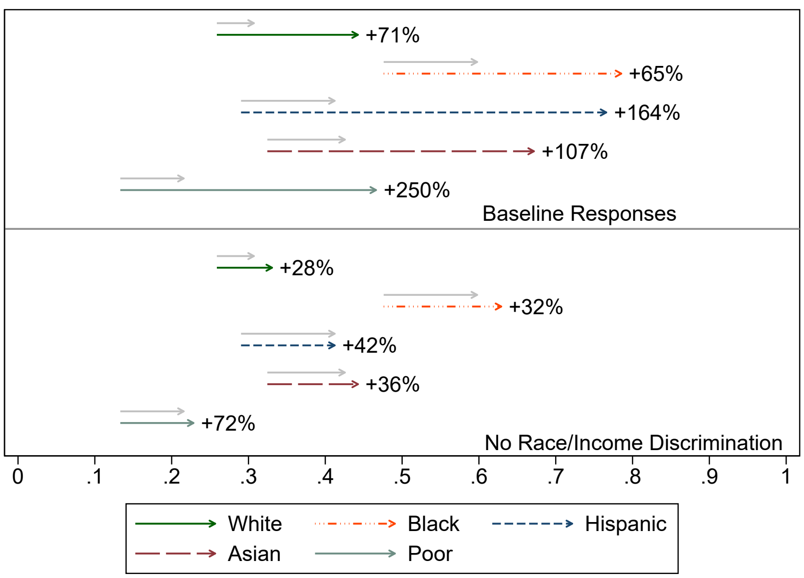
Notes: Each panel shows $\hat{\beta}_{g,g'}$ for all g' for different values of T, the lag of the Instrumental Variable, $\boldsymbol{s_{jt-T}}$. In all panels of this figure g represents poor Whites, Blacks, Hispanics and Asians. We set $T' = 12$ in all cases. For comparison, we keep the sample constant in all cases, so the first 37 months of the sample are not used in the estimation of $\boldsymbol{\beta}$ for all values of T.

Figure 15: Number of Neighborhoods In Flux (Simulated)



Notes: Figure shows the number of neighborhoods with at least one, two, five or ten moves that change their socioeconomic composition (out of a total of 224 neighborhoods). Simulation begins in November 2004 and continues for 200 years.

Figure 16: Equilibrium Changes in Race and Income Segregation - No Moving Costs Ever (Simulated)



Notes: For each panel representing a different counterfactual, the arrows represent changes in simulated Dissimilarity Indices for households of each race and income from November 2004 to sorting equilibrium in the absence of exogenous shocks. Numbers in parentheses correspond to the relative change in dissimilarity. Gray arrows correspond to the baseline changes shown in Figure 3. A Black dissimilarity index of, say, 0.60, means that 60% of Black homeowners would have to be relocated in order to generate an equal distribution of Blacks across all Bay Area neighborhoods.