A Unified Empirical Framework to Study Neighborhood Segregation

Gregorio Caetano and Vikram Maheshri*

February 2024

Abstract

We propose a new framework that incorporates the endogenous feedback loop at the core of the seminal Schelling (1969) model of segregation into a dynamic model of neighborhood choice and use it to study the forces that shaped racial and income segregation in the San Francisco Bay area from 1990 to 2004. Our framework allows us to assess the relative importance in the short- and long-run of various sorting mechanisms that generate segregation – endogenous discriminatory sorting (both taste-based and statistical) on the basis of the socioeconomic composition of neighbors, and sorting on the basis of exogenous changes to neighborhood amenities – along with the frictions that mediate these types of sorting – moving costs and uncertainty. Identification of households' responses to the socioeconomic composition of neighbors is facilitated by novel instrumental variables that exploit the logic of a dynamic choice model with frictions. Discriminatory sorting on the basis of race and income is important, but frictions mitigate its impact on segregation. Because sorting on the basis of other neighborhood amenities shapes segregation, there is scope for place-based desegregation policies. Because of frictions, there is also scope for desegregation policies based on the reallocation of households, particularly in the short-run.

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

1 Introduction

Neighborhood demographics are often in a state of flux. In Figure 1, we present the evolution of the socioeconomic compositions of two San Francisco Bay Area neighborhoods over a fifteen year period. As suggested by these selected neighborhoods, there is heterogeneity in the trends of the race and income compositions across neighborhoods. What explains these trends? In his seminal work, Schelling (1969) proposed a concise answer to this question: the composition of neighborhoods may change due to the presence of information frictions and discrimination.[1] If, for instance, Hispanic households prefer Hispanic neighbors relative to non-Hispanic ones, then an increase in the Hispanic share of a neighborhood might trigger a best response by other Hispanic households to enter the neighborhood, which could in turn trigger further inflows of Hispanics in the future, and so on. Because households do not coordinate on their best responses, this endogenous process may unfold over several periods and generate the observed serial correlation in socioeconomic composition we see in West Richmond.

Meanwhile, a rich parallel literature on residential choice has developed to study sorting on the basis of local amenities (including the socioeconomic compositions of neighborhoods).[2] A common assumption in this literature is that neighborhoods are observed in steady state, i.e., in the absence of future amenity shocks, the demographic compositions of the neighborhoods will not change. This leaves no room for the endogenous mechanism discussed above, so the trends shown in Figure 1 would be attributed to serially correlated, exogenous changes in the amenities of these neighborhoods. For instance, we would conclude that some West Richmond amenity that Hispanics disproportionately like has gradually increased over the sample period in some manner outside of the model of residential choice.

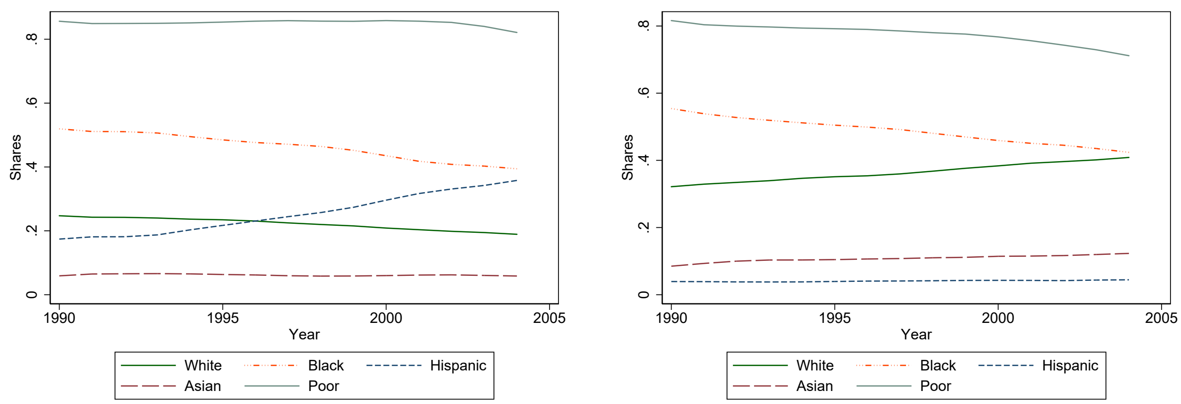
This paper studies the determinants of segregation within a new framework that bridges these two literatures by incorporating into empirical models of residential choice the endogenous feedback loop that fuels the dynamics suggested by Schelling (1969). This is particularly useful to assess the impacts of policies, as this endogenous feedback loop generates dynamic treatment effects: a change in the socioeconomic composition of a neighborhood today may lead to further changes in the future even if no other

[1]Except where explicitly noted, we use the term discrimination to encompass both taste-based and statistical discrimination on the basis of the race and income of neighboring households.

[2]See, for example, McGuire (1974); Epple, Filimon and Romer (1984); Kiel and Zabel (1996); Epple and Sieg (1999); Bayer, McMillan and Rueben (2004*a*,*b*); Bayer and Timmins (2005, 2007); Bayer, Ferreira and McMillan (2007); Wong (2013); Bayer et al. (2016); Caetano (2019).

actions are taken later on. As a result, the short-run effects of such policies may be very different from their long-run effects (Cholli and Durlauf (2022)).

Figure 1: Socioeconomic Composition of Selected San Francisco Bay Area Neighborhoods Over Time, 1990-2004



(a) West Richmond

(b) West Emeryville

Source: See Section 2.

We modify certain standard assumptions in the residential choice literature to accommodate the endogenous Schelling mechanism. This results in two main departures from this literature. First, we exclude as covariates any neighborhood characteristics that may be post-determined from the socioeconomic compositions of neighborhoods. This includes neighborhood prices and other observed amenities (e.g., crime, air quality). This parsimonious approach to specify the choice model is in stark contrast to most of the residential sorting literature, which includes as many observable neighborhood amenities as possible as controls. In a dynamic setting such as ours, including such post-determined covariates generates interpretation issues since these covariates would shut off a portion of the endogenous mechanism driving segregation. Indeed, changes in the socioeconomic compositions of neighborhoods likely affect the levels of prices and other amenities, which in turn may trigger further changes in the socioeconomic compositions of neighborhoods, and so on.[3]

[3]In the appendix, we show in detail how additional data on amenities that are caused by socioeconomic compositions of neighborhoods may be included in the framework. However, this requires further identification assumptions to enable a mediation analysis of segregation dynamics. We then implement this extended framework with price data; while this allows us to consider some additional counterfactuals, we show that all results of our main, parsimonious analysis replicate.

The second main departure from the standard residential sorting literature concerns expectations. Specifically, we relax assumptions about households' expectations at the time they make their residential choices. Because these expectations impact the choices that households make, which in turn trigger changes in the future expectations of the socioeconomic compositions of neighborhoods, restrictions on how expectations are formed may influence the strength of this feedback loop. Indeed, information frictions attenuate the ability of discrimination to translate into segregation because of *ex ante* uncertainty: each household may not individually respond as strongly to the socioeconomic compositions of neighborhoods because they are unable to perfectly forecast the sorting decisions of others or the extent to which amenities will change in the future due to changes in the socioeconomic compositions (leading to a coordination problem among households). This friction is reinforced by the fact that it is costly to move and undo an *ex ante* optimal choice that was made with an expectation that did not get realized.

Identifying this endogenous feedback loop requires us to estimate the causal response of the current period socioeconomic composition of a neighborhood to its socioeconomic composition in the past. For that, we need to isolate households' responses to changes in the compositions of neighborhoods from their responses to other neighborhood amenities, including unobserved ones. To this end, we propose a new instrumental variable (IV) that follows the logic of a dynamic choice model with frictions. Our identification strategy relies on the assumption that information from the more distant past (e.g., two years ago) does not directly affect valuations of neighborhoods today conditional on valuations in the interim (e.g., in the past year). This translates into isolating the component of a neighborhood's socioeconomic composition that is due to *mismatched households*, i.e., those who currently reside in their neighborhood for reasons that are no longer relevant to new inflows; although they made optimal choices in the past given their expectations at that time, these households are now stranded in their homes because of moving costs in spite of that neighborhood having become less attractive to them in the meantime. We perform a detailed sensitivity analysis, including a Monte Carlo study, which together support the validity of the identification strategy in our application.

We analyze a monthly data set of residential sales (Bayer et al. (2016)) across 224 neighborhoods in the San Francisco Bay Area from 1990-2004 that allows us to observe the heterogeneous sorting of eight socioeconomic groups over time: rich and poor Whites, Blacks, Hispanics and Asians. Our empirical framework combines estimation of a dynamic model of neighborhood choice with a simulation procedure that allows us to isolate specific determinants of segregation (and their interactions) in a series of counterfactuals. Our model can be summarized as follows: First, households form expectations about the characteristics of all potential neighborhoods including the (future) proportions of households of every race and income level. Based on these expectations, they decide if they should move and, if so, to which neighborhood they will move. Households of different socioeconomic groups sort heterogeneously on the basis of all neighborhood characteristics, including unobserved amenities. All of this sorting is mediated by two frictions: moving costs and uncertainty. We estimate two key sets of parameters for each of the eight socioeconomic groups of households: their moving costs and their responses to different types of neighbors. Following Bayer et al. (2016), we identify moving costs from the decisions of households who chose to move instead of staying in their current houses. We identify the responses of households to their neighbors with the novel IV strategy discussed above.

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), and different levels of moving costs. We explicitly consider the fact that sorting today affects the choices of households next month, which in turn may affect the choices of households in two months, etc., repeating this endogenous feedback loop indefinitely until a new steady state is reached. This dynamic process is also allowed to spill over across neighborhoods, so we avoid a partial equilibrium analysis as typically seen in dynamic models of segregation (e.g., Becker and Murphy (2000)). By simulating this entire dynamic re-sorting process, we can uncover the resulting trajectories of neighborhoods under these different counterfactuals. A comparison of trajectories across counterfactuals allow us to identify the relative roles of each factor in explaining segregation in both the short-run and the long-run.

We summarize the main results of our empirical analysis in four key findings: (1) Discriminatory sorting is widespread, complex and robust. Households tend to respond positively, though to different degrees, to neighbors of the same race and income. There is also substantial heterogeneity in the responses to neighbors of other types, some of which are not reciprocated. (2) Racial and income discrimination have large impacts on segregation in the long-run. However, discrimination along one dimension (e.g., race) has limited impact on segregation in the other dimension (e.g., income). (3) Moving costs and uncertainty complement each other to substantially mitigate segregation levels. They do so by disproportionately reducing the intensity of discriminatory sorting, and as a result, segregation changes due to discrimination accumulate very slowly over the span of decades. Because of this, there is more scope for other neighborhood amenities to play an important role in explaining segregation. (4) A desegregation policy that equalized the socioeconomic compositions of neighborhoods would be met with intense re-sorting, as most households would be highly mismatched to the amenities of their new neighborhoods. However, even though long-run, steady state levels of segregation would not change dramatically, the glacial pace of re-sorting would still imply large policy effects on segregation over the first few decades.

Relevant Literature

Our paper bridges two distinct but related literatures on residential choice and segregation. It also contributes to a growing literature on the causes of residential 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.[4] 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, Ferreira and McMillan (2007); Bayer, Keohane and Timmins (2009); Ferreira (2010); Bayer and McMillan (2012); Ringo (2013); Bayer et al. (2016); Caetano (2019)). Bayer and Timmins (2005) study the existence and uniqueness of equilibrium in sorting models with endogenous amenities such as the demographic composition of a neighborhood; 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.

[4]Kuminoff, Smith and Timmins (2013) provide a comprehensive review of the growing literature on neighborhood sorting.

Following this literature, we employ a discrete choice framework to relate household choices to their preferences, expectations and constraints. As noted, we make two key departures from this literature. First, we specify a minimal set of observed neighborhood characteristics – the socioeconomic composition of the neighborhood – in order to study the sorting patterns that lead to socioeconomic segregation. This allows us to avoid imposing restrictions on the causal mechanisms through which the socioeconomic composition in the present may affect the socioeconomic composition in the future. Second, we relax assumptions on households' expectations when residential decisions are made. This renders our approach compatible with residential choices being observed along a trajectory towards steady state via the endogenous mechanism described above (as opposed to assuming that all choices are already observed in steady state). In doing so, we take 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 (e.g., 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 uncovering the value of amenities, as is typical in these studies). Moreover, the IV approach that we develop is novel, and it follows from the logic of a dynamic model of neighborhood choice with frictions. These IVs can be created with no additional data requirements, and they can also be used to identify price responses as shown in Appendix B.

Dynamic 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 sort on the basis of the characteristics of their neighbors.[5] 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 a 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 explicitly models a friction, namely myopic expectations, to ensure that neighborhoods gradually evolve toward a steady state.

Subsequent theoretical papers have embedded this intuition into a more standard economic framework (e.g., Becker and Murphy (2000); Bayer and Timmins (2005)),

[5]See Durlauf (2004) for a summary of this literature.

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, 2023)). Banzhaf and Walsh (2013) discuss the role of other amenities in generating segregation under no moving costs. Caetano and Maheshri (2017) and Caetano and Maheshri (2023) study school segregation in a framework that embeds the key insight of Schelling (1969) and discuss how policies may have completely different effects on segregation in the short- and the long-run because of the endogenous feedback loop. In this paper, we generalize and extend that framework in three directions. First, we analyze segregation along multiple dimensions simultaneously, studying both racial and income segregation. Second, we relax 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.

Causes of Residential Segregation

There is a large empirical literature in the social sciences on the causes of Black-White residential segregation that focuses on the deep historical roots of this inequality. Massey and Denton (1993) build on the observations of Myrdal (1944) and articulate numerous causes of racial segregation in the United States, including explicit housing policies in the early twentieth century in both Northern and Southern cities, prejudice, and more modern structural biases in the real estate market (e.g., unequal mortgage access and discrimination by realtors). Indeed, a number of studies have found causal impacts on neighborhood segregation of specific institutional features such as historical zoning regulations and covenants (Shertzer, Twinam and Walsh (2021)), redlining (Aaronson, Hartley and Mazumder (2020)) and stringent credit standards (Ouazad and Rancière (2016)). Cutler, Glaeser and Vigdor (1999) and Boustan (2013) conclude that collective action on the part of Whites in developing both legal and extra-legal institutions explained much of segregation in the mid 20th century, though by the end of the century, individual action in the form of decentralized sorting decisions was the main driver of urban segregation.

In this paper, we do not explicitly model each of these factors separately. Rather, we draw inspiration from Schelling (1969) and differentiate mechanisms of segregation at a higher level between those that generate endogenous feedback (and hence dynamics) and the ones that do not. In that sense, this paper complements this literature since our framework allows for dynamic treatment effects that are important to assess how specific mechanisms such as those studied in this literature may still play an important role today even if they are no longer active. Indeed, moving costs and uncertainty are key impediments to speedy (de)segregation and can explain why institutions that have in many cases been eliminated may still shape current segregation levels. Similarly, our framework helps explain why some desegregation policies might be more successful in the short-run than in the long-run, while other policies might be more successful in the long-run than in the short-run.

We also contribute to this literature by stratifying households into finer groups that allows for heterogeneity across four races/ethnicities (Asian, Black, Hispanic and White) and income levels. This more disaggregated analysis is particularly relevant to a study of contemporary segregation in the United States as Asian and Hispanic households have grown more prevalent. We find that the role of demand-driven discrimination on segregation levels is more complex than previously understood. Some groups discriminate against other groups, but this is not always reciprocated. While discrimination unambiguously increases segregation in the long-run, its quantitative impact on segregation is asymmetric and depends on the socioeconomic group of reference. Importantly, on a shorter time scale demand-driven discrimination has a modest and sometimes non-monotonic impact on segregation since it is mediated by frictions, the initial allocation of households to neighborhoods, and the heterogeneous substitutability of neighborhoods across different socioeconomic groups.

The rest of the paper proceeds as follows. In Section 2, we describe our data. In Section 3, we present an empirical model of neighborhood segregation, articulate the specific assumptions required for identification, and discuss the estimation and simulation of this model. In Section 4, we present our results and consider different counterfactuals in order to assess the importance of various determinants of segregation. Finally, we offer some concluding remarks in Section 5. In the appendix, we extend our framework to explicitly incorporate other endogenous amenities, apply this extended framework with neighborhood prices, replicate our main results, and explore new counterfactuals. We also provide robustness checks, technical details and a Monte Carlo study.

2 Data

We construct a monthly panel 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 are divided 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.[6]

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).[7] For expositional simplicity,

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

[7]We obtain the race and income of the original stock of households as of 1990 from the 1990 Census. From 1990 onward, 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.

we refer to Hispanics as a race rather than an ethnicity, and the other three racial groups include only non-Hispanic households.[8] For each race-income group g, neighborhood j and month t, we observe the total number of homeowners, the total numbers of homeowners who moved into a new house, and the total number of homeowners who stayed in the same home since last month.[9] We also observe the total number of households of each group who chose to exit the Bay Area homeownership market in each month.[10] Finally, we compute monthly neighborhood prices by averaging the sales prices of all transactions observed in the HMDA data.

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

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 for each of the eight socioeconomic groups defined by race and income and summarize it in Table 1.[11] We choose this widely used measure of segregation because it is easy to interpret. For instance, a rich 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 for them to be distributed uniformly across all Bay Area neighborhoods (i.e., to ensure that the

[8]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 adjust the population numbers accordingly.

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

[10]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.

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

share of rich Whites was the same in all 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.

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.

3 Empirical Framework

We first present a dynamic model of neighborhood choice and propose novel instrumental variables to identify key parameters of the model (Section 3.1). We then describe how the parameters of the model are estimated (Section 3.2) and used to study segregation dynamics via simulation (Section 3.3). We conclude with a detailed discussion of the interpretation of model parameters (Section 3.4).

3.1 A Dynamic Model of Neighborhood Choice

A city is divided into J neighborhoods. At the beginning of each period t, household i observes state variable $\boldsymbol{b_{it}}$ as well as where they are currently located, j^*_{it}. They use this information to form expectations of their value of residing in each neighborhood, and then they choose where to reside in order to maximize their expected utility. Formally, household i faces the dynamic optimization problem

$$\max_{j_{i\tau}\in\mathbb{J}} \mathbb{E}\left[\sum_{\tau=t}^{\mathcal{H}} \delta^{\tau-t} \cdot \mathbb{E}[u_{ij_{i\tau}\tau}|\boldsymbol{b_{i\tau}}]\,|j^*_{it}, \boldsymbol{b_{it}}\right], \tag{1}$$

where $j_{i\tau}$ and $\boldsymbol{b_{i\tau}}$ are the choice and state variables of household i in period τ respectively, $u(\cdot)$ is their flow indirect utility function, $\mathcal{H}$ is their time horizon, and δ is their inter-temporal discount factor. (Hereafter, we refer to all vectors and matrices in bold type.) $\mathbb{J}$ is each household's choice set, which includes moving to one of the $j \in \{1, \ldots, J\}$ neighborhoods, remaining in their current house ($j = J + 1$), or leaving the city entirely ($J = 0$).[12]

$\mathbb{E}[u_{ij\tau}|\boldsymbol{b_{i\tau}}]$ corresponds to the flow-utility that household i *expects* to obtain (just prior to making their choice in τ) if they choose neighborhood j, with *state variable* $\boldsymbol{b_{i\tau}}$ denoting the information set used to form that expectation. Our notation differs slightly from the standard literature to make clear the role of expectations in neighborhood sorting. Although $\mathbb{E}[u_{ij\tau}|\boldsymbol{b_{i\tau}}]$ is typically written as $u(j, \boldsymbol{b_{i\tau}})$, we want to make explicit the fact that the flow-utilities are not observed before decisions are made.[13] This allows for an important source of *ex ante* uncertainty over what the resulting socioeconomic composition of neighborhoods will be after all households have made their choices, since households do not generally coordinate on their decisions. Note that so far we have not imposed any restrictions on the structural primitives of the expectation of flow-utility since $\boldsymbol{b_{i\tau}}$ is allowed to be unobserved by the econometrician and may vary freely across

[12] Here we borrow notation for the choice set from Bayer et al. (2016); as in their paper, we only observe data on homeowners, so in our application, $j = 0$ also includes the outside option of renting within the city.

[13] To make our notation otherwise comparable with the standard literature, we follow the same notation as Aguirregabiria and Mira (2010)'s well-known survey of the dynamic discrete choice literature.

i and τ.

We define the value function as $V(\boldsymbol{b_{it}}) = \max_{j\in\mathbb{J}} \mathbb{E}[v_{ijt}|\boldsymbol{b_{it}}]$, where v_{ijt} refers to the cumulative utility of household i for choosing neighborhood j in t. The expectation of this cumulative utility given $\boldsymbol{b_{it}}$ is written as

$$\mathbb{E}[v_{ijt}|\boldsymbol{b_{it}}] = \mathbb{E}[u_{ijt}|\boldsymbol{b_{it}}] + \int \delta \cdot V(\boldsymbol{b_{it+1}})\, dF_b(\boldsymbol{b_{it+1}}\,|j, \boldsymbol{b_{it}}). \tag{2}$$

$F_b(\boldsymbol{b_{it+1}}\,|j, \boldsymbol{b_{it}})$ is the expected distribution of the state variable in $t+1$ conditional on the choice and the state variable from t. Next, we decompose $\boldsymbol{b_{it}}$ into two components, an idiosyncratic and a residual one, and make assumptions on the idiosyncratic term.

Assumption 1. (Additive Separability, Logit Error, Conditional Independence)

1. $\mathbb{E}[u_{ijt}|\boldsymbol{b_{it}}] = \mathbb{E}[u_{ijt}|\boldsymbol{x_{it}}] + \epsilon_{ijt}$ *where* ϵ_{ijt} *is the* j*th element of* $\boldsymbol{\epsilon_{it}}$.
2. ϵ_{ijt} *is i.i.d. extreme value type I.*
3. $F_x(\boldsymbol{x_{it+1}}\,|j, \boldsymbol{x_{it}}, \boldsymbol{\epsilon_{it}}) = F_x(\boldsymbol{x_{it+1}}\,|j, \boldsymbol{x_{it}})$ *where* $F_x(\cdot)$ *is the cumulative density function of* x.

This is a standard assumption in the dynamic discrete choice literature (e.g., Aguirregabiria and Mira (2010)), except for one key departure: in our model we do not assume that $\boldsymbol{x_{it}}$ is observable or estimable by the econometrician (see Remark 2 below). We will impose further assumptions on $\boldsymbol{x_{it}}$ below. Assumption 1 implies

$$\mathbb{E}[v_{ijt}|\boldsymbol{b_{it}}] = \underbrace{\mathbb{E}[u_{ijt}|\boldsymbol{x_{it}}] + \int \delta \cdot \overline{V}(\boldsymbol{x_{it+1}})\, f_{\boldsymbol{x}}(\boldsymbol{x_{it+1}}\,|j, \boldsymbol{x_{it}})}_{\mathbb{E}[v_{ijt}|\boldsymbol{x_{it}}]} + \epsilon_{ijt}, \tag{3}$$

where $\overline{V}(\cdot)$ is the *integrated value function.*[14]

At the beginning of period t, households observe the state variable $\boldsymbol{b_{it}}$, form expectations and then choose (1) whether or not to move, and, upon deciding to move they choose (2) an option in $\mathbb{J} = \{0, \ldots, J\}$. We classify households on the basis of race and income into G different demographic groups indexed by $g = g_i$ (where g_i denotes the group to which household i belongs), and impose the following restriction:

[14]This function, defined as $\overline{V}(\boldsymbol{x_{it}}) = \int V(\boldsymbol{x_{it}}, \boldsymbol{\epsilon_{it}})\, dG_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon_{it}})$, is the unique solution to the integrated Bellman equation $\overline{V}(\boldsymbol{x_{it}}) = \int \max_{j\in\mathbb{J}} \Big(\mathbb{E}[u_{ijt}|\boldsymbol{x_{it}}] + \epsilon_{ijt} + \delta \cdot \sum_{\boldsymbol{x_{it+1}}} \overline{V}(\boldsymbol{x_{it+1}})\, f_x(\boldsymbol{x_{it+1}}\,|j, \boldsymbol{x_{it}})\Big) dG_\epsilon(\boldsymbol{\epsilon_{it}})$, where $G_\epsilon(\boldsymbol{\epsilon_{it}})$ is the extreme value type I cumulative density function.

Assumption 2. *Let* $\mathbf{1}(\cdot)$ *be the indicator function and* $\boldsymbol{x_{it}} = (j_{it-1}, \boldsymbol{x_{g_it}})$*. Then*

$$\mathbb{E}[v_{ijt}|\boldsymbol{x_{it}}] = \mathbf{1}\left(j \in \{0,\ldots,J\}\right) \cdot \left(\mathbb{E}\left[v_{g_ijt}|\boldsymbol{x_{g_it}}\right] - \phi_{g_i}\right) + \mathbf{1}\left(j = J+1\right) \cdot \mathbb{E}\left[v_{g_ijt}|\boldsymbol{x_{g_it}}\right] \quad (4)$$

Assumption 2 states that different households of the same group are allowed to differ from each other only by their previous choice, j_{it-1}, and their idiosyncratic error, $\boldsymbol{\epsilon_{it}}$.[15] It also implicitly states that moving costs do not vary within group or depend on the neighborhoods of origin (j_{it-1}) and destination (j_{it}). These assumptions follow from Bayer et al. (2016) and are needed due to data limitations since we only observe variables at the group-neighborhood-month level.[16]

The term $\mathbb{E}\left[v_{gjt}|\boldsymbol{x_{gt}}\right]$ is the moving-cost-free component of $\mathbb{E}[v_{ijt}|\boldsymbol{x_{it}}]$ for households of group g. Denoting $v^e_{gjt} = \mathbb{E}\left[v_{gjt}|\boldsymbol{x_{gt}}\right]$, $\boldsymbol{s^e_{jt}} = \mathbb{E}[\boldsymbol{s_{jt}}|\boldsymbol{x_{gt}}]$ and $\xi^e_{gjt} = \mathbb{E}[\xi_{gjt}|\boldsymbol{x_{gt}}]$, we decompose:

$$v^e_{gjt} = \boldsymbol{\beta'_g s^e_{jt}} + \xi^e_{gjt}, \quad (5)$$

where $\boldsymbol{s_{jt}}$ is a vector of observed demographic shares representing the socioeconomic composition of each neighborhood and period, and ξ^e_{gjt} encompasses all other neighborhood amenities that are valued by households. It is important to note that equation (5) is written from the perspective of the household, *not* the econometrician, as the three variables reflect the expectations of households when they make their choice. Even if $\boldsymbol{x_{gt}}$ is not observed, v^e_{gjt} can be identified from choice data using standard approaches.[17] However, both variables on the right-hand-side are unobserved to the econometrician. In particular, while $\boldsymbol{s_{jt}}$ is observed, $\boldsymbol{s^e_{jt}}$ is unobserved.

Next, we impose assumptions that allow us to identify $\boldsymbol{\beta_g}$, the effect of the expected socioeconomic composition on the expected cumulative value of the neighborhood. We rewrite equation (5) based on observed quantities as

[15] Note that j_{it-1}, which is a component of $\boldsymbol{x_{it}}$, implicitly appears in equation (4): the last term of equation (4) is equivalent to $\mathbf{1}\left(j = j_{it-1}\right) \cdot \mathbb{E}\left[v_{gj_{it-1}t}|\boldsymbol{x_{gt}}\right]$.

[16] One can in principle allow for heterogeneous parameters across individuals of the same group by observables such as wealth or the neighborhood of origin (e.g., Bayer et al. (2016)), but doing so is infeasible in our application because there are not enough households of certain types inside the same group. Because the observed distribution of residential choices varies greatly by socioeconomic group, some of the heterogeneity in moving costs by neighborhood of origin may be incorporated into our estimates of ϕ_g.

[17] This point is obvious, since $\boldsymbol{x_{gt}}$, whatever it may be, is the information set that the econometrician assumes was used to make choices as observed in the data. This point is typically left implicit in discrete choice models since it is often not made explicit that the left-hand-side of the equation involves an expectation conditional on a generic information set.

$$v^e_{gjt} = \boldsymbol{\beta_g'} \boldsymbol{s_{jt}} + \xi_{gjt}, \tag{6}$$

where $\xi_{gjt} = \xi^e_{gjt} + \boldsymbol{\beta_g'}(\boldsymbol{s^e_{jt}} - \boldsymbol{s_{jt}})$. This is the version of equation (5) from the perspective of the econometrician since v^e_{gjt} can be identified (see below) and $\boldsymbol{s_{jt}}$ is observed. Throughout the remainder of the paper, we refer to ξ_{gjt} as simply "other amenities", although in reality it encompasses both other amenities (ξ^e_{gjt}) as well as any forecast error of $\boldsymbol{s_{jt}}$ using information set $\boldsymbol{x_{gjt}}$ (e.g., from differences in the information sets of different types of households).

Instrumental Variables

In order to identify $\boldsymbol{\beta_g}$, we impose a restriction on the error term ξ_{gjt}.

Assumption 3. *IV Validity. For all g and t, there exists some $T > 0$ such that*

$$\mathbb{E}\left[\xi_{gjt} \,\middle|\, \boldsymbol{v^e_{jt-T}}, \boldsymbol{s_{jt-(T+1)}}\right] = \mathbb{E}\left[\xi_{gjt} \,\middle|\, \boldsymbol{v^e_{jt-T}}\right] \tag{7}$$

Assumption 3 states that sorting in t on the basis of other amenities (ξ_{gjt}) does not use information that was used in the distant past (in the form of $\boldsymbol{s_{jt-(T+1)}}$) unless it is already embedded in valuations from a more recent period ($\boldsymbol{v^e_{jt-T}} = (v^e_{1jt-T}, ..., v^e_{Gjt-T})$). Note that this assumption still allows households to use information from the past to form their expectation of ξ_{jt}. It simply requires that any information from the distant past ($t-(T+1)$ or before) that informs sorting on the basis of amenities in t must have been used by households that sorted more recently, in $t - T$.

Assumption 3 suggests a straightforward strategy to estimate $\boldsymbol{\beta}$ by instrumental variables: we can use $\boldsymbol{s_{jt-(T+1)}}$ as an instrument for $\boldsymbol{s_{jt}}$ in equation (6), provided that we control for $\boldsymbol{v^e_{jt-T}} = (v^e_{1jt-T}, ..., v^e_{Gjt-T})$. This strategy can only be implemented if the instrument is relevant, or more formally:

Assumption 4. *IV Relevance. For all g and t, and for some value of T that satisfies Assumption 3,* $\mathbb{E}\left[\boldsymbol{s_{jt}} \,\middle|\, \boldsymbol{v^e_{jt-T}}, \boldsymbol{s_{jt-(T+1)}}\right] \neq \mathbb{E}\left[\boldsymbol{s_{jt}} \,\middle|\, \boldsymbol{v^e_{jt-T}}\right]$.

Assumption 4 states that $\boldsymbol{s_{jt-(T+1)}}$ and $\boldsymbol{s_{jt}}$ are correlated to each other even conditional on $\boldsymbol{v^e_{jt-T}}$. This is the case because forecast errors and moving costs imply that some households will likely be mismatched to their neighborhood at any given point in time. To see this, note that because households lack perfect foresight, some households residing in a neighborhood as of $t-(T+1)$ would have sorted there because of information that turned out to be orthogonal to $\boldsymbol{v^e_{jt-T}}$ (i.e., the information that led those households to sort to that neighborhood turned out to be irrelevant to later inflows in $t-T$). These households now reside in a less-than-ideal neighborhood in t (they are mismatched), yet many of them remain there because moving is costly, so they still contribute to the socioeconomic composition $\boldsymbol{s_{jt}}$.

We formalize this intuition with the choice model above. Consider a generic period in the distant past, $\tau \leq t-(T+1)$. Let $\boldsymbol{\omega_{gj\tau}} = \boldsymbol{x_{g\tau}} - \mathbb{E}\left[\boldsymbol{x_{g\tau}} \middle| \boldsymbol{v^e_{jt-T}}\right]$. Note that $\boldsymbol{v^e_{jt-T}}$ is influenced by information from periods *after* τ, so $\boldsymbol{\omega_{gj\tau}}$ corresponds to the portion of the information in τ that was later found out to be wrong given the information that was available to households in the subsequent period $(t-T)$. Indeed, some households who sorted in τ may now live in less-than-ideal neighborhoods precisely because of $\boldsymbol{\omega_{gj\tau}}$. These households chose their neighborhood using information that turned out to be irrelevant for future inflows (e.g., because their expectations went unrealized). Some of these households will not have "fixed" their *ex post* mistake by t since their level of dissatisfaction does not exceed the cost of moving. They are precisely the mismatched households described above. Because the demographics of these mismatched households contribute to $\boldsymbol{s_{jt}}$, the erroneous information that they used $(\boldsymbol{\omega_{gj\tau}})$ contributes to $\boldsymbol{s_{jt}}$. However, since $\boldsymbol{\omega_{gj\tau}}$ turned out to be irrelevant to future inflows, it is not used to sort on the basis of ξ_{gjt} by Assumption 3. Note that even though we do not observe $\boldsymbol{x_{g\tau}}$ (and hence $\omega_{gj\tau}$), we can still isolate the variation in $\boldsymbol{s_{jt}}$ stemming from $\omega_{gj\tau}$ for all $\tau \leq t-(T+1)$ and for all g by using the component of $s_{t-(T+1)}$ that is orthogonal to $v^e_{1jt-T}, ..., v^e_{Gjt-T}$ as an IV.

In a scenario without uncertainty, $\boldsymbol{\omega_{gj\tau}} = \boldsymbol{0}$: households would perfectly anticipate the evolution of neighborhoods, and hence no information would ever turn out to be wrong *ex post*. As a result the IVs would be uncorrelated to $\boldsymbol{s_{jt}}$ (i.e., it would not be relevant). However, this uncertainty alone is insufficient to guarantee IV relevance. Without moving costs, households could immediately and freely re-optimize in the face of information that turned out to be wrong *ex post*, so $\boldsymbol{\omega_{gj\tau}}$ would not last more than one period (i.e., $\omega_{gj\tau}$ and s_{jt} would not be correlated to each other for $t > \tau + 1$). It is precisely these two frictions – uncertainty over future neighborhood characteristics and moving costs – that jointly imply IV relevance. This is of course a testable assumption, and we show in Section 4 that it is satisfied.

Remark 1. The exclusion restriction on $\boldsymbol{\xi}$ (Assumption 3) reflects the idea that households in t do not use past information (from before $t-T$) in a more sophisticated manner than households in $t-T$. This is a restriction on the *relative* level of sophistication, not on the *absolute* level, so it is consistent with many formulations of expectations 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 households in t form their expectations with information from the last τ periods, and households in $t-T$ form their expectation with information from the last τ' periods. Then Assumption 3 is compatible with any values of τ and τ' provided that $\tau \leq \tau' + T$. In particular, households in t are allowed to be weakly more sophisticated than those in $t-T$ (i.e., $\tau \geq \tau'$ is allowed).[18]

Remark 2. Standard dynamic discrete choice approaches often parametrically specify the transition probability $f_{\boldsymbol{x}}(\boldsymbol{x_{it+1}}|j, \boldsymbol{x_{it}})$ from equation (3) and assume $\boldsymbol{x}$ is observed or estimable by the econometrician.[19] We want to avoid such assumptions in our context. Because $\boldsymbol{x_{it}} = (j_{it-1}, \boldsymbol{x_{g_i t}})$, and $\boldsymbol{x_{gt}}$ determines the expected compositions of neighborhoods, such assumptions would restrict how neighborhood segregation would evolve over time. Because observed past choices reflect past information sets, we instead relate the information set in $\boldsymbol{x_{gt}}$ with past information sets in $\{\boldsymbol{x_{g't-T}} \, \forall g'\}$, allowing us to state the validity assumptions without the need to explicitly state what must be included or excluded in $\boldsymbol{x_{gt}}$. By not connecting $\boldsymbol{x_{gt}}$ to what econometricians observe, we allow, for instance, the time horizon ($\mathcal{H}$) and the inter-temporal discount factor

[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 setup, there is a distinct asymmetry between the outcome variable (v^e_{gjt}) and the main explanatory variable ($\boldsymbol{s_{jt}}$): while v^e_{gjt} reflects the decisions of those who are choosing a new neighborhood in t, $\boldsymbol{s_{jt}}$ reflects the decisions of many other households as well (e.g., households who made past choices), which may have been mediated by moving costs and different information sets. We exploit this asymmetry to build an identification strategy that relates the information used by households of one group in t with the information used by all past decision makers. Thus, we need only restrict the information used by households in t *relative* to the information used by households in the past. In contrast, identification in the production function literature exploits *absolute* restrictions in the information sets of decision makers (e.g., firms). See Ackerberg (2020) for an illuminating discussion of the identifying assumption made in that literature. Our IV approach is also very different from the shift-share IV approach (e.g., Bartik (1991)). Although both IVs use past shares, ours assumes exogeneity of them only conditional on the valuations in period $t-T$, whereas shift-share IVs assume exogeneity of them unconditionally as discussed by Goldsmith-Pinkham, Sorkin and Swift (2020). See also Almagro and Domínguez-Iino (2022), Li (2023) and Davis, Gregory and Hartley (2023) for recently proposed IVs based on the shift-share approach.

[19] See, Aguirregabiria and Mira (2010) for a survey of the literature.

(δ) to vary across groups, and neither needs to be observed or identified. No further restriction on information sets is required beyond Assumptions 1, 2 and 3.

Remark 3. In principle, one could enrich the specification of the cumulative utility in equation (6) by including prices and/or other neighborhood amenities on the right hand side. This does not come without cost since these variables may be caused by the socioeconomic compositions of neighborhoods, so including them would require a mediation analysis to isolate all channels of causation. In Appendix B, we include price as an additional amenity, and extend the framework to conduct the mediation analysis, obtaining additional empirical results.

3.2 Estimation

Our estimation strategy unfolds in two stages: we first estimate v^e_{gjt} and ϕ_g for all g, j and t (stage 1) and then we estimate $\boldsymbol{\beta_g}$ for all g (stage 2). Before describing our strategy in more detail, it is useful to state the only two pieces of data that are required: (1) Population counts of each group in each neighborhood in each period, which we denote as N_{gjt}. From this, we can derive the socioeconomic composition of residents, the vector $\boldsymbol{s_{jt}}$ whose gth element represents the share of group g:

$$s_{gjt} = \frac{N_{gjt}}{\sum_{g'} N_{g'jt}} \tag{8}$$

(2) The total number of inflows of each group into each of the J neighborhoods, which we denote as I_{gjt}. Note that I_{gjt} cannot be expressed solely in terms of N_{gjt} and N_{gjt-1}, since it also embeds new information about the number of households who chose to remain in the same house from $t-1$ to t.

Stage 1: Estimation of v^e_{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^e_{gjt}. Having decided to move, household i solves the following optimization problem:

$$\max_{j \in \{0,\ldots,J\}} v^e_{g_i jt} - \phi_{g_i} + \epsilon_{ijt} \tag{9}$$

Following Assumption 1, the choice-specific probabilities are

$$
\begin{aligned}
P\left(j_{it}=j \mid j \notin \{J+1\}, j_{it-1}\right) &= \frac{\exp\left(v^e_{g_ijt}-\phi_g\right)}{\sum_{j'=0}^{J}\exp\left(v^e_{g_ij't}-\phi_g\right)} \\
&= \frac{\exp\left(v^e_{g_ijt}\right)}{\sum_{j'=0}^{J}\exp\left(v^e_{g_ij't}\right)}
\end{aligned} \tag{10}
$$

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 $\hat{v}^e_{gjt}$ for $j \in \{0, \ldots J\}$ as

$$
\hat{v}^e_{gjt} = \log\left(I_{gjt}\right) - \log\left(I_{g0t}\right). \tag{11}
$$

Next, we consider the choice of whether or not 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 $J+1$ (not moving) is

$$
\begin{aligned}
P\left(j_{it}=J+1 \mid j_{it-1}=j\right) &= P\left(v^e_{g_ijt}+\epsilon_{iJ+1t} > v^e_{g_ij't}-\phi_{g_i}+\epsilon_{ij't}\ ,\ \forall j' \mid j_{it-1}=j\right) \\
&= \frac{\exp\left(v^e_{g_ijt}\right)}{\sum_{j'=0}^{J}\exp\left(v^e_{g_ij't}-\phi_{g_i}\right)+\exp\left(v^e_{g_ijt}\right)}
\end{aligned} \tag{12}
$$

where the first line must hold for all $j' = 0, \ldots, J$, and the second line follows from the logit formula (Assumption 1.2). The data analog to $P\left(j_{it}=J+1 \mid j_{it-1}=j\right)$ is simply $\frac{N_{g_ijt}-I_{g_ijt}}{N_{g_ijt-1}}$, 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 (12) yields the J empirical moment conditions

$$
h_j\left(\phi_g; \hat{\boldsymbol{v}}^{\boldsymbol{e}}_{\boldsymbol{gt}}\right) = \frac{N_{gjt}-I_{gjt}}{N_{gjt-1}} - \frac{\exp\left(\hat{v}^e_{gjt}\right)}{\sum_{j'=0}^{J}\exp\left(\hat{v}^e_{gj't}-\phi_g\right)+\exp\left(\hat{v}^e_{gjt}\right)} \tag{13}
$$

By plugging our estimates of $\hat{v}^e_{gjt}$ from equation (11) into equation (12), we can estimate ϕ_g by GMM using the moment conditions in (13). In Appendix C, we report results from Monte Carlo simulations based on the model from Section 3.1 that show that $\hat{\boldsymbol{\phi}}$ is a consistent estimator of $\boldsymbol{\phi}$.

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

We rewrite equation (6) based on the observed quantities as[20]

$$\hat{v}^e_{gjt} = \boldsymbol{\beta'_g s_{jt}} + \boldsymbol{\gamma'_g \hat{v}^e_{jt-T}} + \underbrace{\xi_{gjt} + \hat{v}^e_{gjt} - v^e_{gjt} - \boldsymbol{\gamma'_g \hat{v}^e_{jt-T}}}_{\text{error}_{gjt}}, \tag{14}$$

We estimate this equation via Two Stage Least Squares using $\boldsymbol{s_{jt-(T+1)}}$ as an IV for $\boldsymbol{s_{jt}}$, controlling for $\boldsymbol{\hat{v}^e_{jt-T}}$. Based on Assumption 1, $\hat{v}^e_{gjt}$ converges to v^e_{gjt}, and $\boldsymbol{\hat{v}^e_{jt-T}}$ converges to $\boldsymbol{v^e_{jt-T}}$. Moreover, based on Assumption 3, $\boldsymbol{s_{jt-(T+1)}}$ is uncorrelated to ξ_{gjt}, conditional on $\boldsymbol{v^e_{jt-T}}$. In Appendix C, we report results from Monte Carlo simulations based on the model from Section 3.1 that show that $\hat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}$.

3.3 Simulation of Segregation Dynamics

We simulate the dynamics of segregation by considering different counterfactual initial values of the elements of equations (4) and (5). Specifically, we simulate how the demographic composition of each neighborhood will evolve given a counterfactual initial state $\tilde{\boldsymbol{s}}^0 = (\tilde{\boldsymbol{s}}^{\boldsymbol{0}}_{\boldsymbol{1}}, ..., \tilde{\boldsymbol{s}}^{\boldsymbol{0}}_{\boldsymbol{J}})$, counterfactual vector of the initial distribution of households $\tilde{\boldsymbol{N}}^{\boldsymbol{0}}$, whose (g, j)th element is $\tilde{N}^0_{gj}$, counterfactual vector of response parameters $\tilde{\boldsymbol{\beta}}$, counterfactual vector of moving costs $\tilde{\boldsymbol{\phi}}$, and counterfactual matrix of the initial values of other neighborhood amenities $\tilde{\boldsymbol{\xi}}^{\boldsymbol{0}}$, whose (g, j)th element is $\tilde{\xi}^0_{gj}$. Each counterfactual is defined by the tuple $\left(\tilde{\boldsymbol{s}}^{\boldsymbol{0}}, \tilde{\boldsymbol{N}}^{\boldsymbol{0}}, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{\xi}}^{\boldsymbol{0}}\right)$ where tildes are used to denote that each of these are counterfactual objects that we can manipulate in order to analyze the effects of different forces on segregation.

Below we define the recursive function $\boldsymbol{s}\left(\cdot; \tilde{\boldsymbol{N}}^{\boldsymbol{0}}, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{\xi}}^{\boldsymbol{0}}\right)$ such that $\boldsymbol{s}\left(\tilde{\boldsymbol{s}}^\tau; \tilde{\boldsymbol{N}}^{\boldsymbol{0}}, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{\xi}}^{\boldsymbol{0}}\right) = \tilde{\boldsymbol{s}}^{\boldsymbol{\tau+1}}$ for each $\tau \geq 0$, which characterizes the evolution of neighborhood compositions. To simplify the exposition, we write $\boldsymbol{s}(\tilde{\boldsymbol{s}}^\tau)$ with the understanding that $\left(\tilde{\boldsymbol{N}}^{\boldsymbol{0}}, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{\xi}}^{\boldsymbol{0}}\right)$ is held constant. For $\tau \geq 0$, the $(g, j)^{\text{th}}$ element of the function $\boldsymbol{s}\left(\tilde{\boldsymbol{s}}^{\boldsymbol{\tau}}\right)$, $\tilde{s}^{\tau+1}_{gj}$, is calculated as

[20] In principle, we could specify $\hat{v}^e_{gjt}$ as a more flexible function of $\boldsymbol{s_{jt}}$ in equation (6). We did so by specifying it with a cubic spline with two knots, but our results were broadly unchanged. This is consistent with the findings of Caetano and Maheshri (2017).

$$\tilde{s}_{gj}^{\tau+1}\left(\tilde{\boldsymbol{s}}^{\boldsymbol{\tau}}\right)=\frac{N_{gj}\left(\tilde{\boldsymbol{s}}^{\boldsymbol{\tau}}\right)}{\sum_{g'} N_{g'j}\left(\tilde{\boldsymbol{s}}^{\boldsymbol{\tau}}\right)} \tag{15}$$

where

$$\begin{aligned} N_{gj}\left(\tilde{\boldsymbol{s}}^{\boldsymbol{\tau}}\right)= \quad & N_{gj}\left(\tilde{\boldsymbol{s}}^{\boldsymbol{\tau-1}}\right)\times\left(\frac{\exp\left(v_{gj}^{e}\left(\tilde{s}^{\tau}\right)\right)}{\sum_{j'=1}^{J}\exp\left(v_{gj'}^{e}\left(\tilde{s}^{\tau}\right)-\tilde{\phi}_{g}\right)+\exp\left(v_{gj}^{e}\left(\tilde{s}^{\tau}\right)\right)}\right)+ \\ & \sum_{k=1}^{J} N_{gk}\left(\tilde{\boldsymbol{s}}^{\boldsymbol{\tau-1}}\right)\times\left(\frac{\exp\left(v_{gj}^{e}\left(\tilde{s}^{\tau}\right)-\tilde{\phi}_{g}\right)}{\sum_{j'=1}^{J}\exp\left(v_{gj'}^{e}\left(\tilde{s}^{\tau}\right)-\tilde{\phi}_{g}\right)+\exp\left(v_{gk}^{e}\left(\tilde{s}^{\tau}\right)\right)}\right) \end{aligned} \tag{16}$$

with $v_{g0}^{e}\left(\tilde{\boldsymbol{s}}^{\boldsymbol{\tau}}\right)$ normalized to zero, and

$$v_{gj}^{e}\left(\tilde{\boldsymbol{s}}^{\boldsymbol{\tau}}\right)=\tilde{\boldsymbol{\beta}}_{\boldsymbol{g}}^{\boldsymbol{\prime}}\tilde{\boldsymbol{s}}_{\boldsymbol{j}}^{\boldsymbol{\tau}}+\tilde{\xi}_{gj}^{0}. \tag{17}$$

Equation (15), which is analogous to (8), simply defines the group shares. Equation (16) is based on the logit formula (e.g., see equation (12)). The first term on the right-hand side corresponds to the simulated number of households who resided in neighborhood j in the previous period and remained in their house, incurring no moving costs. The second term represents the simulated number of households who resided in neighborhood k in the previous period and then moved to neighborhood j next period (households with $k=j$ moved houses within neighborhood j). By implementing this simulation simultaneously for all neighborhoods, we incorporate all endogenous changes in demographics due to sorting that spill over from one neighborhood to another. Throughout the simulation, we keep the total number of households of each group across all neighborhoods in the Bay Area constant at the level implied by $\tilde{\boldsymbol{N}}^{\boldsymbol{0}}$ to ensures that our results do not reflect aggregate demographic changes to the SF Bay area, which are outside of our model.[21]

By repeatedly evaluating $\boldsymbol{s}\left(\cdot\right)$ starting from $\tilde{\boldsymbol{s}}^{\boldsymbol{0}}$, we can use equations (15), (16) and (17) to construct a *simulated trajectory* $\mathbb{T}\left(\tilde{\boldsymbol{s}}^{\boldsymbol{0}}\right)=\{\tilde{\boldsymbol{s}}^{\boldsymbol{0}},\tilde{\boldsymbol{s}}^{1},\tilde{\boldsymbol{s}}^{2}...\}$. We define a *steady state* as follows:

[21] In our simulations, this is achieved by disallowing the current stock of households from switching from an "inside" option to an outside option or vice-versa. In practice, we exclude the outside option of leaving the Bay Area homeowners market by omitting $\exp\left(v_{g0}^{e}-\phi_{j}\right)$ from the denominators of the probabilities in equation (16), and we do not consider households who lived outside the Bay Area as potential inflows.

Definition 1. State $\boldsymbol{s}^{\star}$ is a *steady state* if $\mathbb{T}\left(\tilde{\boldsymbol{s}}^{\mathbf{0}}\right)$ converges to $\boldsymbol{s}^{\star}$ for some $\tilde{\boldsymbol{s}}^{\mathbf{0}}$.[22]

In our framework, there is the possibility of multiple steady states since different initial states $\left(\tilde{\boldsymbol{s}}^{\mathbf{0}}, \tilde{\boldsymbol{N}}^{\mathbf{0}}, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{\xi}}^{\mathbf{0}}\right)$ may converge to different steady states. Each steady state corresponds to how neighborhoods would look in the long-run under a given counterfactual initial condition defined by $\left(\tilde{\boldsymbol{s}}^{\mathbf{0}}, \tilde{\boldsymbol{N}}^{\mathbf{0}}, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{\xi}}^{\mathbf{0}}\right)$, so it can be useful when considering the long-run effects of counterfactual policies on segregation. We could in principle identify all steady states by conducting a grid search of all possible counterfactuals $\left(\tilde{\boldsymbol{s}}^{\mathbf{0}}, \tilde{\boldsymbol{N}}^{\mathbf{0}}, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{\xi}}^{\mathbf{0}}\right)$ and simulating trajectories $\mathbb{T}\left(\tilde{\boldsymbol{s}}^{\mathbf{0}}\right)$ for each of them, but the complexity of such an undertaking is formidable. Even just conducting a grid search on $\tilde{\boldsymbol{s}}$ and fixing all other initial conditions is computationally complex beyond the scope of this paper because of the high dimensionality of $\tilde{\boldsymbol{s}}$. For tractability, we restrict our analysis to a specific set of informative counterfactuals.[23]

Choices of $\left(\tilde{\boldsymbol{s}}^{\mathbf{0}}, \tilde{\boldsymbol{N}}^{\mathbf{0}}, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{\xi}}^{\mathbf{0}}\right)$

As noted above, we must fix initial conditions and parameters before performing any simulation. We discuss these choices here. Let $\mathcal{T}$ correspond to November 2004, the final period of our sample.

1. Initial socioeconomic composition ($\tilde{\boldsymbol{s}}^{\mathbf{0}}$) - Our baseline choice for the initial socioeconomic composition is the observed $\boldsymbol{s}_{\mathcal{T}}$. We also consider a counterfactual in which households of each group are reallocated across neighborhoods so that the socioeconomic compositions of all neighborhoods are identical. This sheds light on the dynamic effects of a policy that fully integrated the Bay Area.

2. Initial distribution of households ($\tilde{\boldsymbol{N}}^{\mathbf{0}}$) - Our baseline choice is the observed $\boldsymbol{N}_{\mathcal{T}}$.

3. Response parameters ($\tilde{\boldsymbol{\beta}}$) - Our baseline choice for the values of the response parameters is their estimated values $\hat{\boldsymbol{\beta}}$. We consider counterfactuals where specific elements of $\tilde{\boldsymbol{\beta}}$ are set to zero (corresponding to scenarios where households are "race-blind", "income-blind", or both).

[22] This notion of "steady state" in this paper has been sometimes referred to as "equilibrium" in the theoretical literature on the dynamics of segregation (Schelling (1969, 1971); Becker and Murphy (2000)). We view "steady state" as a more appropriate term because neighborhoods are understood to be always in Perfect Bayesian Equilibrium in our setup.

[23] In a simpler and more tractable context, Caetano and Maheshri (2017) uses a framework nested inside ours to consider all possible counterfactual initial conditions and report all steady states for each school in Los Angeles. In this paper we consider only a subset of counterfactuals, as the set of possible initial conditions in this general framework is much larger.

4. Moving Costs ($\tilde{\boldsymbol{\phi}}$) - Our baseline choice for the moving cost parameters are their estimated values $\hat{\boldsymbol{\phi}}$. We also consider a counterfactual in which there is a one-time amnesty where $\tilde{\boldsymbol{\phi}} = \mathbf{0}$ for a single period, and one in which moving costs are repeatedly set to zero.

5. Initial amenities ($\tilde{\boldsymbol{\xi}}^{\mathbf{0}}$) - Our baseline choices for the initial unobserved expected value of amenities are their estimated values in the final period of our sample, i.e., $\tilde{\boldsymbol{\xi}}^{\mathbf{0}} = \hat{\boldsymbol{\xi}}_{\boldsymbol{\mathcal{T}}}$, whose (g, j)th element is $\hat{\xi}^0_{gj\mathcal{T}} = \hat{v}^e_{gj\mathcal{T}} - \hat{\boldsymbol{\beta}}'_{\boldsymbol{g}} \boldsymbol{s}_{\boldsymbol{j\mathcal{T}}}$. This ensures that our simulation corresponds to one in which there are no *exogenous* shocks to amenities after November 2004. However, we should note that endogenous changes to amenity levels (which are determined by endogenous variation in $\tilde{\boldsymbol{s}}^{\mathbf{1}}, \tilde{\boldsymbol{s}}^{\mathbf{2}}, \dots$) are included in $\hat{\boldsymbol{\beta}}$. See Section 3.4.

Remark 4. There is an important connection between assumptions on expectations and the endogenous feedback loop. Note that $\boldsymbol{s}\left(\boldsymbol{s}^{\boldsymbol{e}}_{\boldsymbol{t}}\right) = \boldsymbol{s}_{\boldsymbol{t}}$ by construction since the observed choices of each group in t are made when $\tilde{\boldsymbol{s}} = \mathbb{E}[\boldsymbol{s}_{\boldsymbol{t}} | \boldsymbol{x}_{\boldsymbol{gt}}] = \boldsymbol{s}^{\boldsymbol{e}}_{\boldsymbol{t}}$. Thus, if we assume that households are able (and have always been able) to perfectly forecast the compositions (i.e., $\boldsymbol{s}^{\boldsymbol{e}}_{\boldsymbol{t}} = \boldsymbol{s}_{\boldsymbol{t}}$ for all g), then $\boldsymbol{s}\left(\boldsymbol{s}_{\boldsymbol{t}}\right) = \boldsymbol{s}_{\boldsymbol{t}}$; that is, data would always be observed in steady state and no feedback loop would exist. In fact, the specific trajectory of convergence to the steady state is likely affected by expectations, so it is a good idea to avoid strong restrictions to the formation of households' expectations.

3.4 Interpretation of Model Parameters

Interpretation of $\boldsymbol{\beta}$ The coefficient matrix $\boldsymbol{\beta}$ should *not* be interpreted as a preference parameter; it simply captures the various responses of households of different groups to their expectations of the socioeconomic compositions of neighborhoods across all potential channels. Moreover, it does not reveal whether these responses are mediated through changes in flow utilities or the continuation values associated with neighborhood choices. To see this, consider equation (3), and define u^e_{gjt} (flow utility) and CV^e_{gjt} (continuation value from choosing neighborhood j in t) as the averages of $\mathbb{E}\left[u_{ijt} | \boldsymbol{x}_{\boldsymbol{it}}\right]$ and $\int \delta \cdot \overline{V}\left(\boldsymbol{x}_{\boldsymbol{it+1}}\right) f_{\boldsymbol{x}}\left(\boldsymbol{x}_{\boldsymbol{it+1}} \left| j, \boldsymbol{x}_{\boldsymbol{it}}\right.\right)$ across all households of group g respectively. Then we can write $v^e_{gjt} = u^e_{gjt} + CV^e_{gjt}$. For each g and g', $\beta_{g,g'} = \frac{\partial v^e_{gjt}}{\partial s_{g'jt}} = \frac{\partial u^e_{gjt}}{\partial s_{g'jt}} + \frac{\partial CV^e_{gjt}}{\partial s_{g'jt}}$, i.e., $\beta_{g,g'}$ represents the *total* marginal effect of an increase in the expected g' share on the group g valuation of that neighborhood.

We focus on the total effect because it allows us to study segregation dynamics without the additional assumptions required for this decomposition (see Remark 2). As Manski (2004) argues, choice data alone is insufficient to separately identify expectations and preferences. For instance, suppose a neighborhood is expected to increase its poor share, and we observe rich households responding to it by reducing their demand for that neighborhood. From choice data alone, we could not conclude that they responded to prejudice against poor households (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. While this would prevent us from identifying, say, households' willingness to pay to avoid residing close to poor neighbors, it would not restrict us from analyzing how households sort into or out of a neighborhood in response to an increase in the poor share since this is fundamentally related to households' choices and not their preferences *per se*. By imposing the assumptions that would allow us to identify households' preferences for their neighbors' characteristics without actually observing data on households' expectations, we would necessarily restrict how households' form their expectations. This is unwise in our setting, as this would in turn restrict the simulated trajectories that we identify (Remark 4).

Thus, $\boldsymbol{\beta}$ includes any type of discriminatory sorting on the basis of the socioeconomic composition of neighbors, including pure socioeconomic animus (or affinity) and statistical discrimination. It is useful to elaborate on what may constitute statistical discrimination in the context of neighborhood sorting. In the example above, suppose rich households inferred from the poor share in t that the quality of the neighborhood school will decline in the future. A response to that expectation would qualify as statistical discrimination.[24] Thus, $\boldsymbol{\beta}$ includes not only sorting on the basis of changes in socioeconomic compositions *per se* ($\boldsymbol{s_{jt}}$), but also sorting on the basis of expected future endogenous changes in other amenities (e.g., ξ_{jt+1}) due to changes in socioeconomic compositions. We are not aware of any empirical paper that separately identifies these different forms of neighborhood discrimination.

Moreover, $\boldsymbol{\beta}$ may also reflect supply-driven discrimination. For instance, suppose we found that Black households responded positively to an increase in the Black share. This would be possible even if Blacks exhibited no demand-driven discrimination, whether

[24]In a world of complete information, households would not use the neighbors' attributes to predict other amenities in the future, as they would be able to know their values directly. Uncertainty leads them to use such information. See Fang and Moro (2011) for a survey of models of statistical discrimination.

taste-based or statistical. Indeed, the same pattern could alternatively be explained by Black households simply facing obstacles to residing in neighborhoods without Blacks because of discrimination on the part of, say, the mortgage market (e.g., Ladd (1998)) or real estate agents (e.g., Ondrich, Ross and Yinger (2003)). Using the language of Christensen and Timmins (2019), in this example supply-driven discrimination would "steer" Blacks toward Black neighborhoods, which would lead us to find that Blacks respond positively to Black share even if there was no demand-driven discrimination on their part. If we had information about how the choice set of certain groups are more restricted because of such supply-driven discrimination, we could in principle separately identify such effect. Because we do not, we follow the standard approach of assuming all groups have the same choice set, hence $\boldsymbol{\beta}$ incorporates this effect as well.

Summing up, $\boldsymbol{\beta}$ reflects the overall *ability* of households to discriminate, i.e. to sort on the basis of the expected socioeconomic composition of the neighborhood for whatever reason. This ability is affected by both demand and supply considerations. Frictions (such us uncertainty over how neighborhoods will evolve) may restrict or enhance this ability, so they are reflected in $\boldsymbol{\beta}$ as well. The only friction that is not included in $\boldsymbol{\beta}$ is the one that we explicitly model in the paper: (current period) moving costs.

Interpretation of $\boldsymbol{\phi}$ The moving cost parameter, $\boldsymbol{\phi}$, refers to the *current* moving cost of changing homes. That is, although $\boldsymbol{\phi}$ enters separately from $\boldsymbol{\beta}$ in the model, this only applies to moves in period t. If households also anticipated expected costs of future moves (in response to unforeseen changes in neighborhood characteristics), those costs would be loaded into CV_{gjt}. Thus, $\boldsymbol{\beta}$ may also contain components related to the interaction between the anticipated possibility of forecast errors and expected future moving costs. Indeed, households may recognize that they are unable to perfectly predict the future socioeconomic compositions of neighborhoods and any associated effects on other neighborhood amenities, and this may necessitate a future (costly) move.

4 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).[25]

4.1 Estimation Results

In Table 2, we present estimates of the responses to the socioeconomic compositions of neighborhoods ($\boldsymbol{\beta_g}$) along with the moving costs (ϕ_g) for households of each group. The endogenous amenities $\boldsymbol{s_{jt}}$ are instrumented by $\boldsymbol{s_{jt-13}}$ in equation (14).[26]

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 $s_{g'jt}$ 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 rich households respond positively to rich neighbors. Hispanics respond most positively to neighbors of their own race, followed by Asians and Blacks. Own race responses are stronger for poor households than rich households. Interestingly, not all responses are reciprocated: e.g., rich Hispanics respond negatively to Blacks, and poor Hispanics respond positively to Blacks, but Blacks of both income groups show little response to Hispanics. Altogether, these heterogeneous responses may give rise to complex dynamics.

Own race responses range from roughly one third of moving costs to the entirety of moving costs, while cross race responses tend to be smaller. [27] This allows for the possibility 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).

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

[26] For robustness, we also estimated a specification where we included a cubic B-spline of $v_{g'jt-12}$ for all g' with four knots as controls in equation (14) to ensure that we were appropriately controlling for households' valuations of neighborhoods in period $t = T$. Our results were effectively unchanged.

[27] As discussed in Kennan and Walker (2011), household-level moving costs in such discrete choice frameworks can be interpreted as also including ϵ (defined in Assumption 1), so they may vary substantially across households. This can explain why some households would move even with such large gaps between $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$. Thus, moving costs *conditional on moving* are far less prohibitive than the moving cost estimates shown in Table 2. As a robustness check, we allowed for moving costs to vary by both group and year, but we found little heterogeneity over time.

Table 2: Responses to the Socioeconomic Compositions of Neighborhoods

	White		Black		Hispanic		Asian	
	Rich	Poor	Rich	Poor	Rich	Poor	Rich	Poor
Responses to:								
Black Share	-9.47	-7.41	9.41	10.71	-1.08	3.51	-3.67	-1.30
	(0.35)	(0.43)	(0.41)	(0.42)	(0.36)	(0.39)	(0.39)	(0.36)
Hispanic Share	-15.02	-5.93	-0.32	0.24	25.78	28.19	-1.13	4.94
	(0.55)	(0.69)	(0.51)	(0.52)	(0.59)	(0.63)	(0.50)	(0.58)
Asian Share	-4.50	-10.10	-0.87	-2.34	-1.84	-4.02	18.47	21.40
	(0.34)	(0.44)	(0.38)	(0.35)	(0.40)	(0.38)	(0.51)	(0.52)
Poor Share	-4.77	4.74	-5.11	2.01	-8.50	-0.20	-12.15	0.47
	(0.36)	(0.43)	(0.29)	(0.28)	(0.34)	(0.35)	(0.40)	(0.39)
Moving Costs	28.57	28.70	27.44	27.60	28.04	28.16	28.06	27.64
	(0.01)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)
Num. of Observations				147,840				

Notes: The first eight columns of responses are 2SLS estimates of $\boldsymbol{\beta}_g$ from equation (14). We use $s_{g'jt-13}$ for all g' as instrumental variables. White is the omitted racial share and rich is the omitted income share. Moving costs are estimated by GMM (see equation 13). All standard errors clustered by group-month. The p-values for both the Cragg-Donald and the Kleinbergen-Paap weak identification tests are less than 0.001, which imply strong first stages.

Sensitivity Analysis

In Appendix A, we present raw OLS estimates of $\boldsymbol{\beta}$ (Table 3). Our OLS estimates of $\boldsymbol{\beta}$ 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 similarly), all of which would bias 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 $T+1$ (the period corresponding to our IVs) in Figures 8 and 9. Larger values of $T+1$ relax Assumption 3, resulting in an IV that is more likely to be valid. We find that estimates of every element of $\boldsymbol{\beta}$ changes very little for $T+1 = 13, \ldots, 36$.

In Appendix C, we present results from a Monte Carlo exercise to study how our IV performs in practice. We first simulate data using our model of neighborhood choice, and then we implement our IV approach for different values of T and show that our estimator of β performs well.

4.2 Counterfactual Analysis

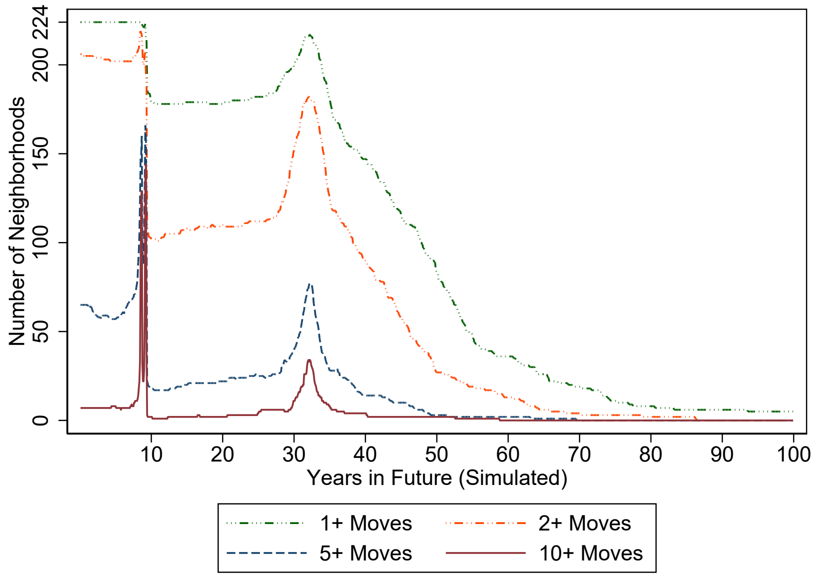
4.2.1 Baseline

We simulate the evolution of the socioeconomic compositions of neighborhoods by setting $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\phi}}$ equal to their estimated values and $\tilde{\boldsymbol{s}}$, and $\tilde{\boldsymbol{N}}$ at their observed November 2004 values. This set of initial conditions and parameters corresponds to our "Baseline" simulation, which can be interpreted as the simulated dynamic trajectory of segregation from November 2004 onward in the absence of future external shocks. 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 household 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 several decades to follow, nearly all neighborhoods are in flux. From this, we conclude that the Bay Area is not observed to be in steady state.[28] Despite substantial moving costs, the amenities of the neighborhoods where households are observed to reside are sufficiently unattractive to some households that most neighborhoods experience turnover. Over time, changes in the socioeconomic compositions of these neighborhoods trigger further sorting, which generates a dynamic feedback loop that changes the relative attractiveness of each neighborhood in different ways across

[28] This also implies that neighborhoods are not observed at "tipping points" since they correspond to an unstable steady state. Hence, small deviations in our simulation due to, say, estimation error, should leave our long-run conclusions effectively unchanged, which we confirmed empirically.

Figure 2: Number of Neighborhoods In Flux (Simulated)



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

different socioeconomic groups. This process is slow and non-monotonic: although it takes 90-100 years for the Bay Area to approximate steady state[29], there are brief episodes of greater churn around 10 years and 33 years after the simulation begins, when gradual changes in a small number of neighborhoods suddenly trigger moves that spill over to a large number of neighborhoods (because households inflow to a small set of neighborhoods from a wide set of other neighborhoods, or vice-versa). This general equilibrium "tipping" that unfolds across a large number of neighborhoods at the same time is a different manifestation of tipping than the notion of tipping in a single neighborhood (e.g., Schelling (1971); Zhang (2009)) and has not been studied to the best of our knowledge.

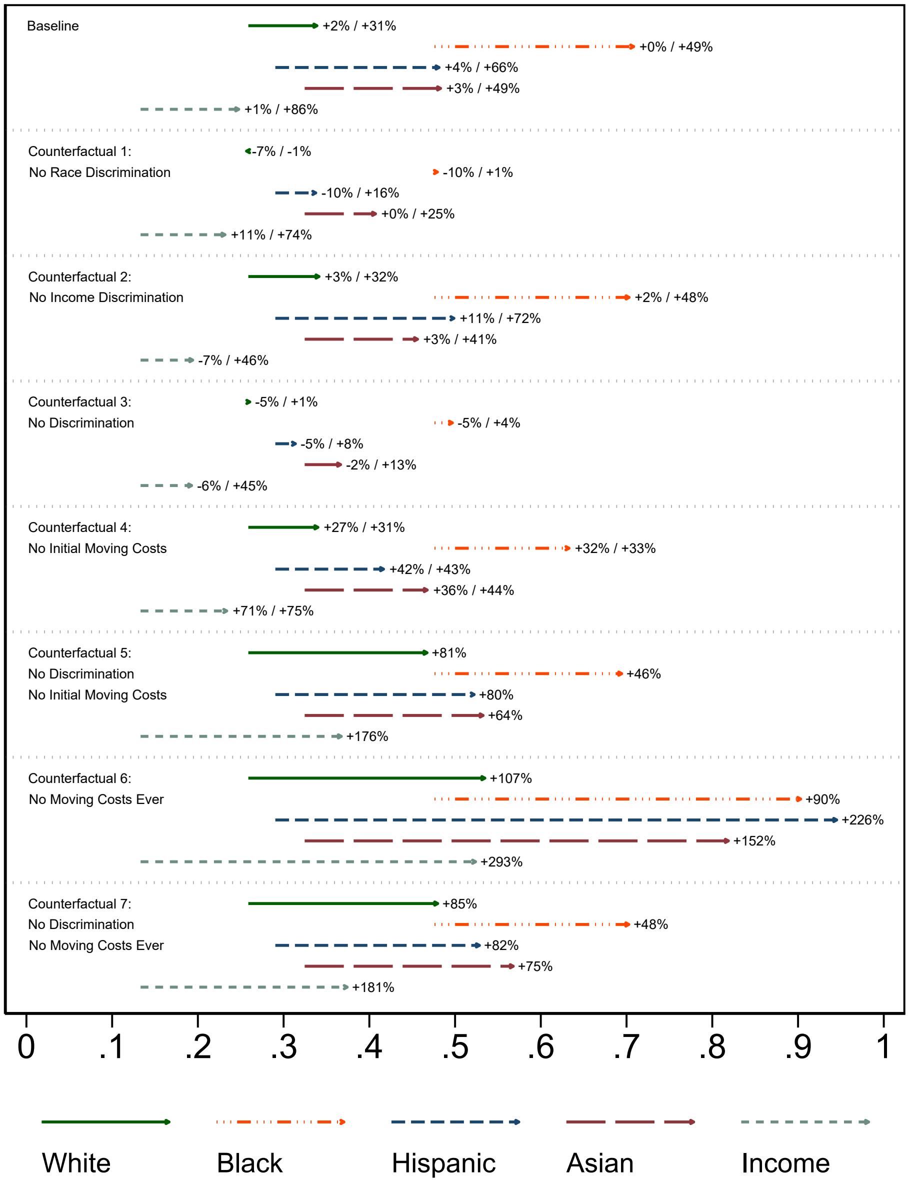
The outcome of this pattern of sorting is a change in the levels of segregation in the Bay Area. We present these results in Figure 3. The top panel corresponds to our baseline simulation, and the remaining nine panels correspond to different counterfactuals that we discuss in more detail in the remainder of this section. In each

[29]It takes 168 years for all Bay Area neighborhoods to experience no net moves.

panel, we present simulated changes in the dissimilarity index for each race (pooling income groups) and for each income group (pooling races) across all Bay Area neighborhoods. In every simulation, the change for each group is presented as an arrow with two numbers corresponding to the medium-run (5 years into the future) and the long-run (steady state) change in the dissimilarity index. The steady state corresponds to the moment where there are no longer any (net) moves into any neighborhoods, which in the baseline counterfactual occurs after 168 years.

As shown in the top panel labeled "Baseline", all races experience modest increases in segregation, though this is a very slow process. White households experience the smallest increase in segregation in both absolute and relative terms: in the medium-run, they are only 2% more segregated, and in the long run they are 31% more segregated. Black households start off more segregated than all other races and remain so throughout the simulation. However, Black segregation is effectively unchanged in the medium run, though it increases by 49% in the long run. Hispanic homeowners experience the largest absolute and relative (4% in the medium-run, 66% in the long-run) increases in segregation, followed by Asians, who experience a 3% increase in segregation in the medium-run, and a 49% increase in the long-run. Although income segregation increases substantially in the long-run, this process unfolds slowly (1% in the medium-run, 86% in the long-run). This relative change is large because of the low level of income segregation at the start of the simulation. Indeed, income segregation remains lower than racial segregation throughout the entire simulation.

Figure 3: Medium- and Long-Run Changes in Race and Income Segregation (Simulated)



Notes: The arrows represent the changes in simulated dissimilarity indices for households of each race/income from November 2004 onward in the absence of shocks to ξ. (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 numbers shown correspond to medium-run (5 years into the future) and long-run (168 years into the future, after neighborhoods have reached steady state) changes in segregation relative to November 2004. Medium-run and long-run changes are identical in counterfactuals 4-7. Details of each counterfactual are presented in Section 4.2.

4.2.2 The Roles of Discriminatory Responses: Race and Income

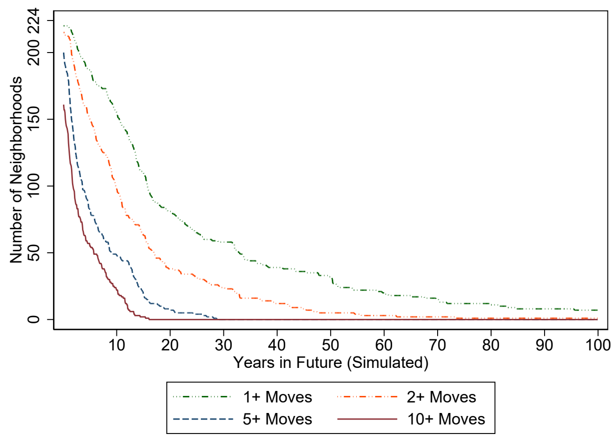
Our estimates of $\hat{\boldsymbol{\beta}}$ (Table 2) reveal systematic discriminatory responses of households of all socioeconomic groups. To isolate their roles in explaining the patterns of segregation presented in Figures 2-3, we consider a series of counterfactuals in which households are either "race-blind", i.e., unresponsive to the racial composition of their neighbors ($\tilde{\beta}_{g,g'} = 0$ for all g and $g' \in \{\text{Black}, \text{Hispanic}, \text{Asian}\}$), "income-blind", i.e., unresponsive to the income composition of their neighbors ($\tilde{\beta}_{g,\text{poor}} = 0$ for all g), or both race- and income-blind ($\tilde{\boldsymbol{\beta}} = \mathbf{0}$). Note that this eliminates all discriminatory mechanisms, both taste-based and statistical. Moreover, as we shut off each type of discriminatory sorting, uncertainty over the racial and/or income composition of neighborhoods can no longer contribute to further sorting. For instance, when we make households race-blind, we not only make them unresponsive to changes in the racial composition of their neighbors, we also eliminate the possibility of *ex post* mismatch because other households sorted in an unexpected way.

We present the simulated changes in segregation levels under each of these counterfactuals in Figure 3 (panels labeled Counterfactual 1-3). In comparison to the baseline results, it is clear that removing racial discrimination has a profound impact on segregation in both the medium and long run (Counterfactual 1). Indeed, segregation is expected to decrease slightly in the medium run for all races except for Asians in the absence of racial discrimination, and any long run increases are modest at best. Our findings of decreases in segregation in the medium-run is evidence that households and neighborhood amenities are initially mismatched. What keeps households of a given race in a particular neighborhood is its racial composition; once this is no longer valued, they are much more likely to move to a different neighborhood with a very different racial composition. Income segregation increases faster in the medium-run, which suggests that eliminating racial discrimination leaves more scope for income discrimination to affect sorting. As this modified feedback loop driven by neighborhood income and other amenities unfolds, households converge to a new steady state with a similar amount of income segregation as in the baseline but much lower racial segregation.

In counterfactual 2, we eliminate income discrimination instead. The results are analogous to counterfactual 1 with one exception. As in counterfactual 1, we find medium-run impacts of eliminating income discrimination on segregation in both racial

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

(a) No Racial or Income Discrimination



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

and income dimensions, and we find segregation in the other (racial) dimension converges to the same level as in the baseline simulation. However, income segregation rises much more in the long run in counterfactual 2 than racial segregation rises in counterfactual 1. Still, this increase in long-run income segregation is only about half of the baseline increase. In counterfactual 3, we eliminate both racial and income discrimination, and our conclusions from counterfactuals 1 and 2 are unchanged, which suggests that there is no meaningful complementarity between sorting on the basis of each demographic dimension.

In Figure 4, we present the number of neighborhoods that are in flux for each period into the future under counterfactual 3. It is evident that discrimination and uncertainty are together responsible for non-monotonic convergence.

Remark 5. Figure 4 shows a gradual dynamic adjustment even in the absence of discrimination (when there is no scope for Schelling-style segregation dynamics). This arises due to moving costs and uncertainty stemming from ϵ_{ijt} shocks to households (all other sources of uncertainty are shut off since ξ is fixed in all simulations and Counterfactual 3 assumes $\tilde{\boldsymbol{\beta}} = 0$). In each period, households receive a new shock that was not previously anticipated (ϵ_{ijt} in equation (9)). This leads some households to become sufficiently mismatched that they choose to move away from their current house because they anticipate a higher utility from a new house (net of moving costs) than their anticipated utility from their current house. These households are the new inflows in the subsequent period, which leads to segregation dynamics. With each move, households become better matched, hence it becomes less likely that an ϵ_{ijt} shock will be sufficiently large to trigger a further move.[30]

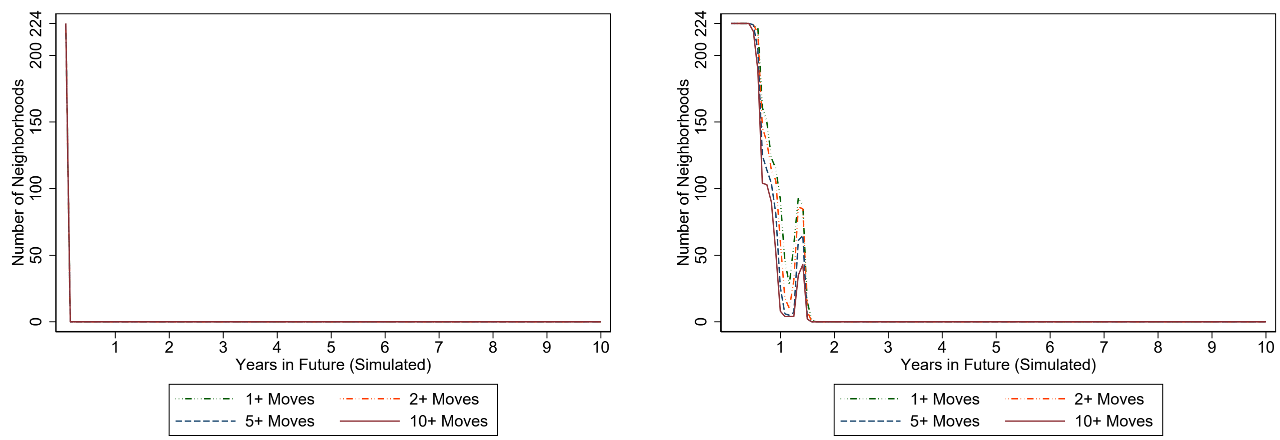
4.2.3 The Role of Moving Costs

The slow declines of Figures 2 and 4 suggest that moving costs play an important role in shaping segregation dynamics. To explore this further, we consider a counterfactual in which all households enjoy a one-time moving-cost amnesty at the beginning of the simulation ($\tilde{\boldsymbol{\phi}}$ is set to zero for the first iteration of the simulation and then set to $\hat{\boldsymbol{\phi}}$ thereafter). As shown in the first panel of Figure 5, the Bay Area converges to a steady state instantaneously. In the first period, the lack of moving costs allows households to eliminate their mismatch (per their *ex ante* expectation in t). However, this does not imply that there is no mismatch in t or in further periods, as forecast errors (due to uncertainty over how other households will act) may lead households to reside in neighborhoods that turn out to be suboptimal. Nevertheless, this mismatch is quite small relative to moving costs, which are restored in future periods and prevent further moves.

[30]To see this, consider a household who is currently (as of period $\tau > \mathcal{T}$, where $\mathcal{T}$ represents November 2004) residing in neighborhood j and anticipates neighborhood j' to be their ideal option (net of moving costs). Their current mismatch is given by $\xi_{j'\mathcal{T}} + \epsilon_{ij'\tau} - (\xi_{j\mathcal{T}} + \epsilon_{ij\tau}) > 0$ (note that under no discrimination $\mathbb{E}[v_{ikt}|\boldsymbol{x_{it}}] = \xi_{kt}$ for all k). When this mismatch is smaller, it is less likely that future i.i.d. realizations of $\epsilon_{ij'\tau}$ and $\epsilon_{ij\tau}$ will generate a move (i.e., $\xi_{j'\mathcal{T}} - \phi_{g_i} + \epsilon_{ij'\tau} - (\xi_{j\mathcal{T}} + \epsilon_{ij\tau}) > 0$) when previous realizations did not generate a move (i.e., $\xi_{j'\mathcal{T}} - \phi_{g_i} + \epsilon_{ij'\tau'} - (\xi_{j\mathcal{T}} + \epsilon_{ij\tau'}) \leq 0$ for $\mathcal{T} \leq \tau' < \tau$). Note that $\epsilon_{ij'\tau} - \epsilon_{ij\tau}$ is drawn from a logit distribution, which is decreasing in the right tail.

Figure 5: 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.

In order to confirm that there is still mismatch after the first period, consider a different counterfactual where we eliminate moving costs by repeatedly setting $\tilde{\boldsymbol{\phi}} = \mathbf{0}$ in each period of the simulation. This exercise is valuable as it allows us to gauge the role of forecast errors in shaping segregation dynamics. To see this, note that if households' forecasts of other households actions were perfectly accurate, then all mismatch would be resolved after the first period since everyone will have had an opportunity to move costlessly to their ideal neighborhood, and perfect forecasts would ensure that these neighborhoods would remain their ideal *ex post* (since there would be no scope for coordination failure). Thus, in the absence of forecast errors, we should not observe further moves if they were granted further moving cost amnesties. We present neighborhood dynamics when moving costs are always zero in the second panel of Figure 5; these findings suggest that households are unable to perfectly forecast the choices of other households. While convergence is still much faster than in the baseline case with moving costs (note that the horizontal axis corresponds only to a 10 year period as opposed to a 100 year period), it does takes longer than one period, owing to the fact that forecast errors trigger further costless moves.[31]

[31]In counterfactual 7 (Figure 3), we permanently set $\tilde{\boldsymbol{\phi}} = \mathbf{0}$ with no discrimination (i.e., we also set $\tilde{\boldsymbol{\beta}} = \mathbf{0}$) and obtain instantaneous convergence to steady state again. This is to be expected,

We explore the interaction between moving costs and discrimination in counterfactuals 4-7 of Figure 3. Relative to the baseline case, counterfactual 4 (which maintains baseline discriminatory responses) leads to higher immediate segregation levels, but they are not necessarily higher in the long-run. In fact, a comparison with counterfactual 5 (no initial moving costs and no discrimination) shows that segregation levels are ultimately lower in the presence of discrimination. Forecast errors help explain this result. Note that since neighborhoods can be thought of as bundles of several characteristics, a household's ideal neighborhood in terms of demographics ($\boldsymbol{s}$) is generally different from their ideal neighborhood in terms of other amenities (ξ). Households consider both characteristics when sorting in counterfactual 4, but they only consider ξ in counterfactual 5. As a result, $\boldsymbol{s}$ triggers more diffuse sorting across households of the same socioeconomic group in counterfactual 4, yielding lower long-run segregation levels than in counterfactual 5. This bundling issue would not arise in the absence of uncertainty since households would effectively coordinate on their ideal neighborhood in terms of ξ. Moving costs, which are reestablished after the first period, prevent *ex post* mismatch (after an initial costless move) from being resolved. This is confirmed by a comparison of counterfactuals 4 and 6: segregation increases much more in the long-run when moving costs are repeatedly eliminated (counterfactual 6) since coordination issues are no longer relevant, and households can coordinate around their ideal neighborhood in terms of ξ. Indeed, note that segregation increases by a nearly identical amount in counterfactuals 5 and 7, suggesting that sorting on the basis of ξ alone does not appreciably generate mismatch after an initial period of costless moving.

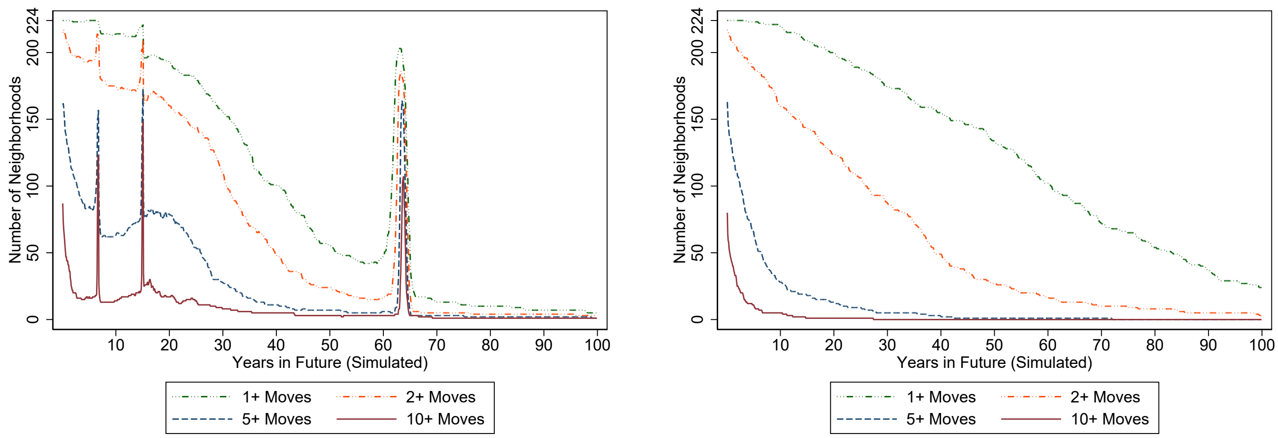
Remark 6. For counterfactual 6 to be sensible, we must make an additional assumption to ensure that $\hat{\boldsymbol{\beta}}$ does not change when future moving costs are set to zero. Recall that since $\hat{\boldsymbol{\beta}}$ was estimated from choice data, it includes the continuation value associated with sorting on the basis of $\boldsymbol{s}$, which generally incorporates the anticipatory utility associated with the possibility that households may someday regret their choice and want to undertake a costly move again. A natural interpretation that rationalizes this assumption is to consider a counterfactual policy that surprises households with additional moving cost amnesties in all periods after the first. Although this policy is unrealistic, it serves as a helpful device to understand the role of uncertainty in explaining segregation.

since we remove all scope for discrimination so forecast errors in the socioeconomic compositions of neighborhoods are no longer relevant and cannot generate coordination failure.

Figure 6: 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 net moves that change their socioeconomic composition (out of a total of 224 neighborhoods) under a different counterfactual. Simulation begins in November 2004.

4.3 The Role of the Initial Allocation of Households

We now consider a hypothetical policy in which households are re-allocated so that all neighborhoods have the exact same initial socioeconomic compositions (but, importantly, other neighborhood amenities ξ are unchanged). The first panel of Figure 6 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 only slightly longer to reach steady state, though there are additional episodes of churn farther out into the future. This could reflect the fact that such a policy leads to misalignment that takes longer to undo because of moving costs. Eliminating discrimination, as in the second panel of Figure 6, ensures convergence is monotonic as before, though it also seems to slow down convergence as expected.

We explore the relationship between initial socioeconomic compositions and segregation in Figure 7 under six counterfactuals. The numbers indicating percentage changes in this Figure are all measured relative to observed segregation levels just prior to the reallocation policy (November 2004). When starting in a fully integrated Bay Area, the segregation level is zero by construction. counterfactual $0'$ shows that this integration policy would reduce segregation in the medium-run relative to the baseline counterfactual, but it would not lead to large relative reductions in segregation in the long-run (and would even increase Hispanic segregation). This can be seen by comparing the ending points of the arrows for counterfactual $0'$ and the baseline. In contrast, when we eliminate discrimination in counterfactual $3'$, segregation of all types would reduce dramatically, even in the long-run. This suggests that desegregating policies might not be effective in the long-run due to discrimination.[32]

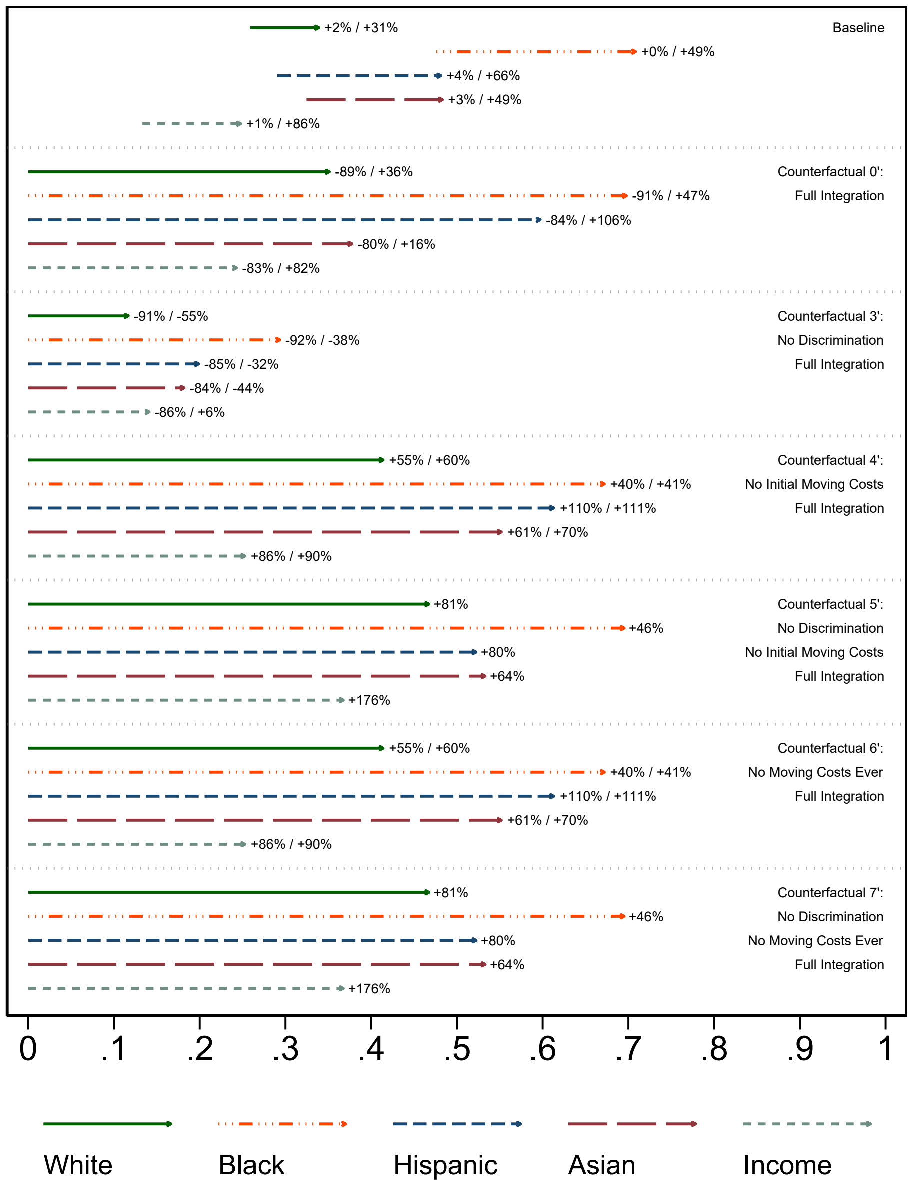
In counterfactuals $4'$-$7'$, moving costs are eliminated either in a one time amnesty or permanently. This rapidly intensifies the levels of segregation in the medium-run relative to counterfactual $0'$, but changes the long-run levels of segregation very little. The stark contrast between counterfactuals $4'$-$7'$ and 4-7 from Section (5) is evidence of the role of uncertainty. To see this, note that without uncertainty over the demographic trajectories of neighborhoods, there should be no difference between counterfactuals 4 and $4'$ or between counterfactuals 6 and $6'$. Indeed, the initial allocation of households should only matter in the absence of uncertainty if moving was costly. However, with uncertainty, the initial allocation matters for a second reason: the demographic distribution implied by the initial allocation affects $\boldsymbol{s}$, and households use that value of $\boldsymbol{s}$ to inform their (discriminatory) sorting. In counterfactuals $4'$ and $6'$ all neighborhoods have identical $\boldsymbol{s}$, which leaves room for sorting only due to ξ; on the contrary, in counterfactuals 4 and 6, households balance discriminatory sorting with non-discriminatory sorting since neighborhoods start with different values of $\boldsymbol{s}$. This also explains why the changes from counterfactuals 5 and 7 are similar to the changes from counterfactuals $5'$ and $7'$, respectively, as in these cases there is no room for discriminatory sorting, so there is no room for a coordination problem due to uncertainty.

5 Conclusion

Neighborhoods constantly evolve: their amenities are not static and their residents are in flux. Theoretical (disequilibrium) models of segregation tend to attribute this evolution to endogenous changes in neighborhood residents arising from discrimination, while

[32]This is consistent with the empirical literature on demographic change associated with school segregation, which Morrill (1989) summarizes as "Over the longer term, the effect of mandatory busing is to ... [foster] polarization between a poorer, minority-dominated central city and richer, white-dominated suburbs."

Figure 7: Steady State Changes in Race and Income Segregation - Full Integration (Simulated)



Notes: The arrows represent the changes in simulated dissimilarity indices for households of each race/income from November 2004 onward in the absence of shocks to ξ. (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 numbers shown correspond to medium-run (5 years into the future) and long-run (after neighborhoods have reached steady state) changes in segregation relative to November 2004. Medium-run and long-run changes are identical in counterfactuals $5'$ and $7'$. Details of each counterfactual are presented in Section 4.3.

disaggregated (equilibrium) models of residential choice tend to attribute this evolution to exogenous changes in other amenities. In this paper, we develop an empirical framework that synthesizes 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, uncertainty, moving costs, and the initial allocation of households across neighborhoods, we explore the underlying forces that drive segregation through counterfactual analyses. A key finding is that while discriminatory sorting is an important factor that contributes to segregation, frictions (moving costs and uncertainty) prevent much desired discriminatory sorting from occurring. This weakens the feedback loop at the core of the Schelling model and leaves more scope for other amenities to explain segregation, especially in the short-run. In a frictionless world, policymakers would have only place-based desegregation policies at their disposal to have lasting effects. However, given the frictions that we find, policymakers might also employ people-based desegregation policies (e.g., nudging or explicitly incentivizing specific people to move to specific neighborhoods) to making lasting impacts. The specifics of an optimal policy would depend on many different contextual features of the choice environment such as the intensity of households' preferences, their expectations, the magnitudes of frictions, and the substitutability of different groups across different neighborhoods. Our framework provides insight into each of these factors and the complex complementarities between them (i.e., the effects of one factor depend on the intensities of other factors).

A well known concern in studies of neighborhood segregation is the lack of high resolution, high frequency data. With better data, our framework could be used to conduct a more detailed empirical analysis in four key directions. First, better demographic data would allow us to classify households into finer socioeconomic groups and study the dynamics of segregation along additional margins. Second, high frequency historical data would allow us to use this framework to understand segregation as a historical phenomenon and also explore the extent to which segregation today is due to historical inequalities, including some that might no longer exist but are still currently relevant. Third, high frequency demographic data on renters would allow for a more complete analysis of the patterns of neighborhood segregation in San Francisco. Fourth, high resolution and high frequency data on other neighborhood amenities (e.g., the supply of neighborhood venues) might allow for an analysis of the effects of policies that target these characteristics on socioeconomic segregation, provided that the additional identification assumptions required for an appropriate mediation analysis are plausible.

Ultimately, we view the main contribution of this paper to be a framework for the empirical analysis of determinants of segregation that can be readily adapted to various contexts subject to data availability. In that sense, the minimal data requirements of our framework should facilitate future research. The use of this framework to study sorting along different demographic dimensions (e.g., race, income, partisanship, education) and in different settings (e.g., neighborhoods, schools, physical venues, social media) could prove valuable to understand important cleavages in our society.

References

Aaronson, Daniel, Daniel A Hartley and Bhashkar Mazumder. 2020. "The Effects of the 1930s HOLC'Redlining'Maps.".

Ackerberg, Daniel A. 2020. "Timing Assumptions and Efficiency: Empirical Evidence in a Production Function Context." *Working Paper* .

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

Almagro, Milena and Tomás Domínguez-Iino. 2022. "Location sorting and endogenous amenities: Evidence from amsterdam." *Available at SSRN 4279562* .

Angrist, J.D. and J.S. Pischke. 2009. *Mostly harmless econometrics: An empiricist's companion*. Princeton Univ Pr.

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.

Bartik, Timothy J. 1991. "Who benefits from state and local economic development policies?".

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 and Robert McMillan. 2012. "Tiebout sorting and neighborhood stratification." *Journal of Public Economics* 96(11-12):1129–1143.

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.

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

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.

Boustan, Leah Platt. 2013. Racial residential segregation in American cities. Technical report National Bureau of Economic Research.

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. 2023. "Explaining Recent Trends in US School Segregation." *Journal of Labor Economics* 41(1):175–203.

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

Cholli, Neil A and Steven N Durlauf. 2022. "Intergenerational mobility.".

Christensen, P. and C. Timmins. 2019. "Sorting or Steering: Experimental Evidence on the Economic Effects of Housing Discrimination." *NBER Working Paper* .

Cutler, David M, Edward L Glaeser and Jacob L Vigdor. 1999. "The rise and decline of the American ghetto." *Journal of political economy* 107(3):455–506.

Davis, D., J. Gregory and D. Hartley. 2023. "Preferences over the Racial Composition of Neighborhoods: Estimates and Implications." *Working Paper* .

Durlauf, Steven N. 2004. "Neighborhood effects." *Handbook of regional and urban economics* 4:2173–2242.

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

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.

Fang, Hanming and Andrea Moro. 2011. "Theories of statistical discrimination and affirmative action: A survey." *Handbook of social economics* 1:133–200.

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.

Goldsmith-Pinkham, Paul, Isaac Sorkin and Henry Swift. 2020. "Bartik instruments: What, when, why, and how." *American Economic Review* 110(8):2586–2624.

Kennan, John and James R Walker. 2011. "The effect of expected income on individual migration decisions." *Econometrica* 79(1):211–251.

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.

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.

Ladd, Helen F. 1998. "Evidence on discrimination in mortgage lending." *Journal of Economic Perspectives* 12(2):41–62.

Li, Nicholas Y. 2023. "Racial Sorting, Restricted Choices, and the Origins of Residential Segregation in U.S. Cities." *mimeo.* .

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

Massey, Douglas and Nancy A Denton. 1993. *American apartheid: Segregation and the making of the underclass*. Harvard university press.

McGuire, Martin. 1974. "Group segregation and optimal jurisdictions." *Journal of Political Economy* 82(1):112–132.

Morrill, Richard L. 1989. "School busing and demographic change." *Urban Geography* 10(4):336–354.

Myrdal, G. 1944. "An American dilemma.".

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

Ondrich, Jan, Stephen Ross and John Yinger. 2003. "Now you see it, now you don't: why do real estate agents withhold available houses from black customers?" *Review of Economics and Statistics* 85(4):854–873.

Ouazad, Amine and Romain Rancière. 2016. "Credit standards and segregation." *Review of Economics and Statistics* 98(5):880–896.

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

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.

Shertzer, Allison, Tate Twinam and Randall P Walsh. 2021. "Zoning and segregation in urban economic history." *Regional Science and Urban Economics* p. 103652.

Wong, Maisy. 2013. "Estimating ethnic preferences using ethnic housing quotas in Singapore." *Review of Economic Studies* 80(3):1178–1214.

Zhang, J. 2009. "Tipping and Residential Segregation: A Unified Schelling Model." *Journal of Regional Science* 51:167–193.

Online Appendix

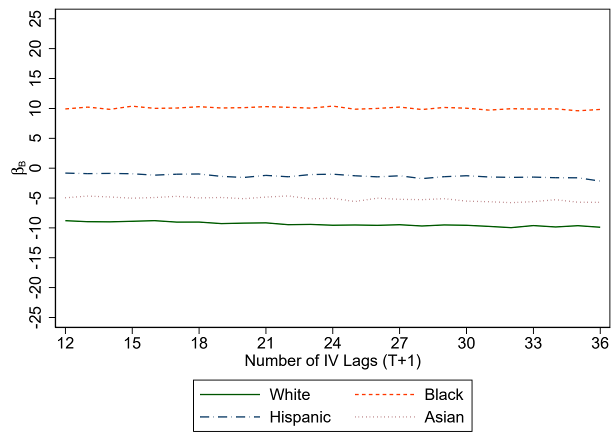
A Tables and Figures

Table 3: OLS Estimates of Responses to the Race and Income Compositions of Neighborhoods ($\boldsymbol{\beta}$)

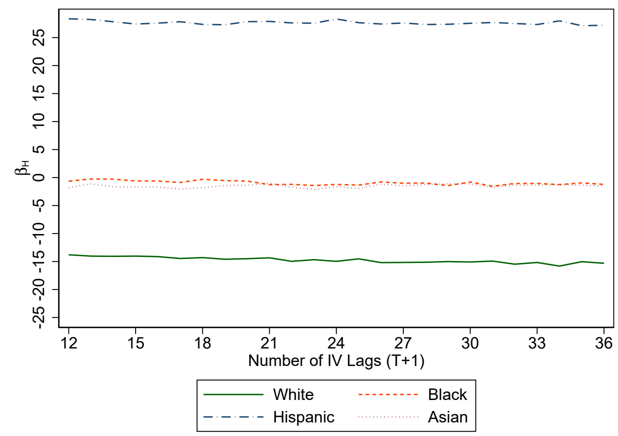
	White		Black		Hispanic		Asian	
	Rich	Poor	Rich	Poor	Rich	Poor	Rich	Poor
Responses to:								
Black Share	-12.25	-7.32	13.47	16.35	-0.81	8.49	-4.81	1.02
	(0.35)	(0.52)	(0.43)	(0.43)	(0.33)	(0.35)	(0.37)	(0.34)
Hispanic Share	-17.69	2.66	8.18	8.86	35.92	43.64	-1.76	16.25
	(0.51)	(0.53)	(0.47)	(0.38)	(0.67)	(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.

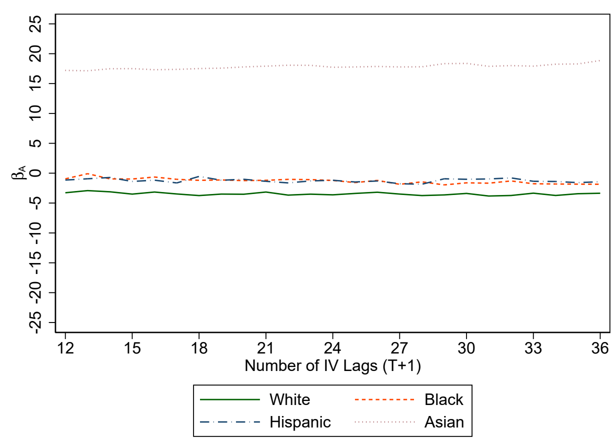
Figure 8: Responses of Rich Households of Different Races to Race and Income Compositions for Different Values of T



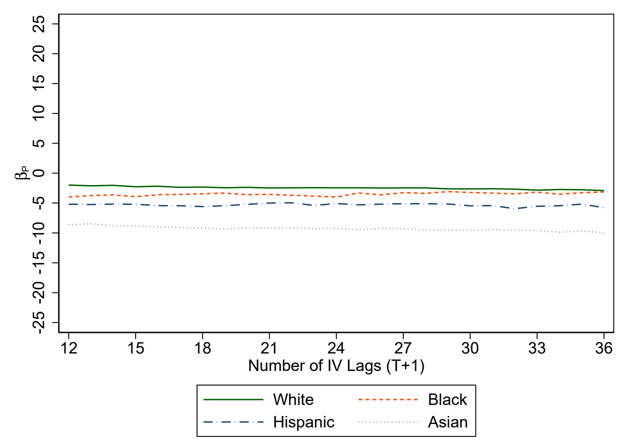
(a) Responses to Black Share



(b) Responses to Hispanic Share



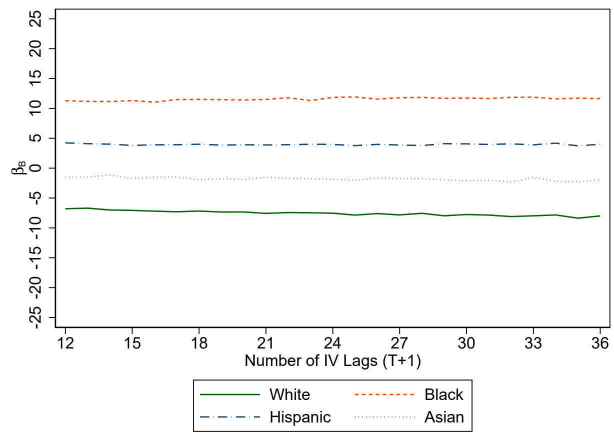
(c) Responses to Asian Share



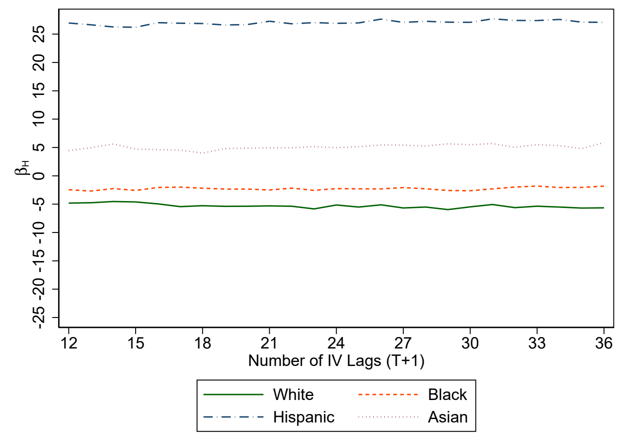
(d) Responses to Poor Share

Notes: Each panel shows $\hat{\beta}_{g,g'}$ for all g' for different values of $T+1$, the lag of the Instrumental Variable, $\boldsymbol{s_{jt-(T+1)}}$. In all panels of this figure g represents poor Whites, Blacks, Hispanics and Asians. We control for $\boldsymbol{v_{jt-T}}$ in all specifications. 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 $\boldsymbol{\beta}$ in any specification.

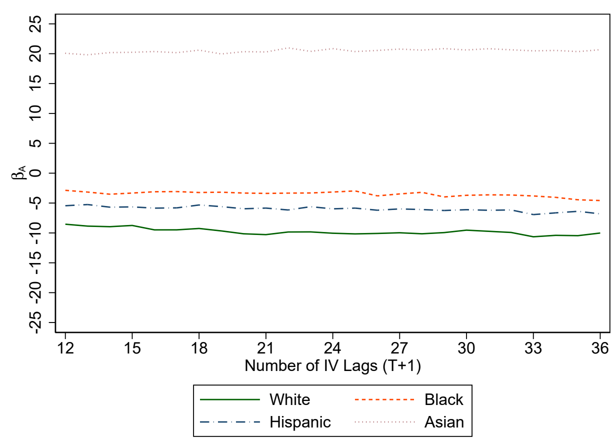
Figure 9: Responses of Poor Households of Different Races to Race and Income Compositions for Different Values of T



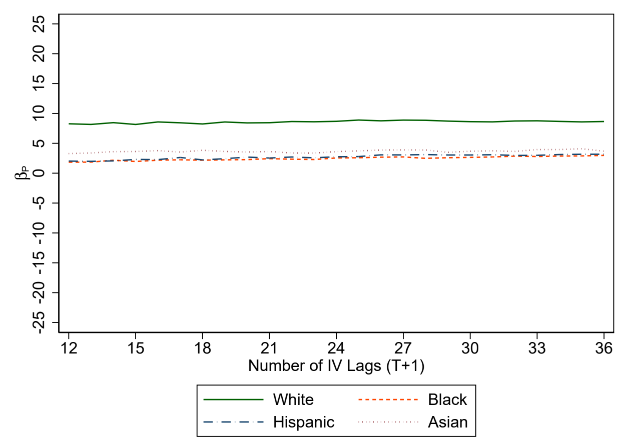
(a) Responses to Black share



(b) Responses to Hispanic Share



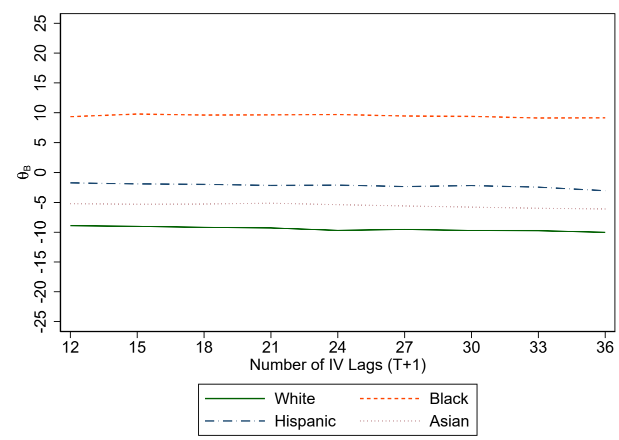
(c) Responses to Asian Share



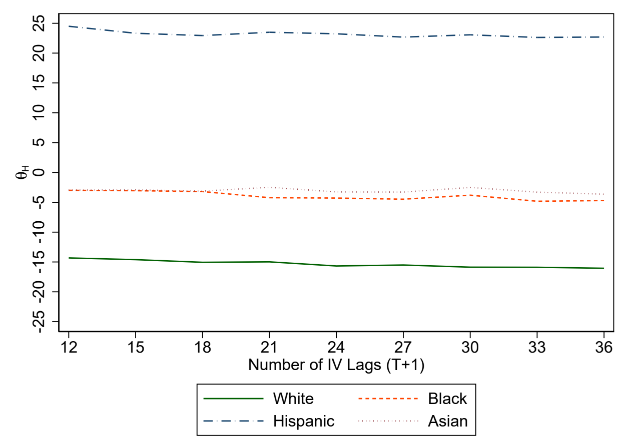
(d) Responses to Poor Share

Notes: Each panel shows $\hat{\beta}_{g,g'}$ for all g' for different values of $T+1$, the lag of the Instrumental Variable, $\boldsymbol{s_{jt-(T+1)}}$. In all panels of this figure g represents poor Whites, Blacks, Hispanics and Asians. We control for $\boldsymbol{v_{jt-T}}$ in all specifications. 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 $\boldsymbol{\beta}$ in any specification.

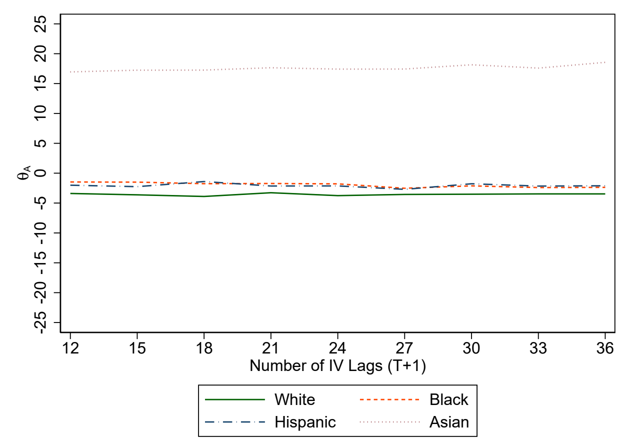
Figure 10: Responses of Rich Households of Different Races to Race and Income Compositions ($\theta_{g,g'}$) for Different Values of T: Price Explicitly Included



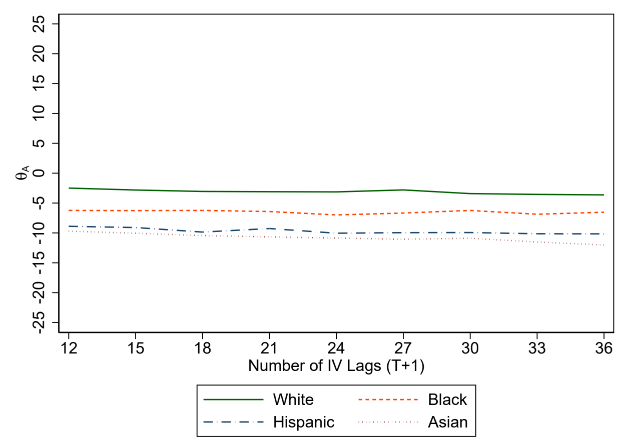
(a) Responses to Black Share



(b) Responses to Hispanic Share



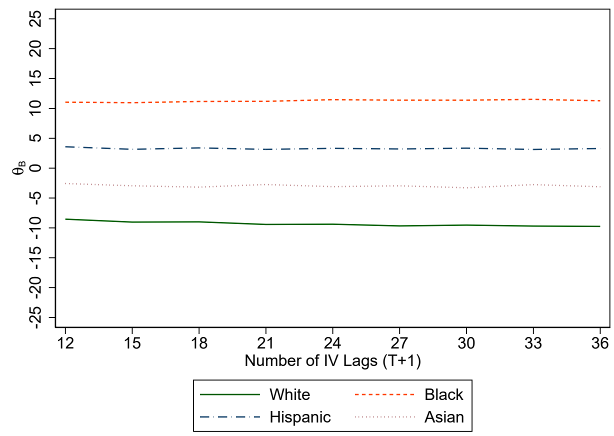
(c) Responses to Asian Share



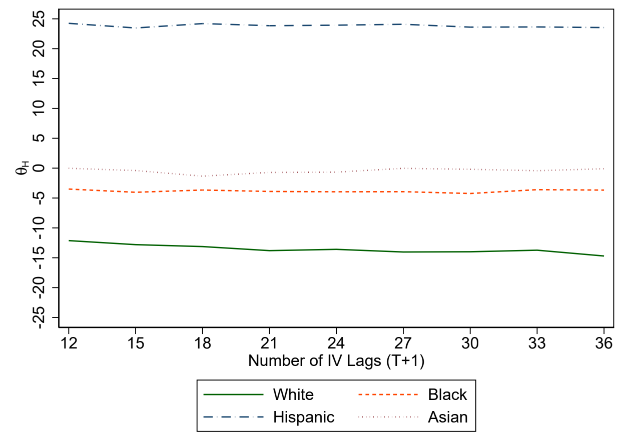
(d) Responses to Poor Share

Notes: Each panel shows $\hat{\theta}_{g,g'}$ for all g' for different values of $T+1$, the lag of the Instrumental Variable, $\boldsymbol{s_{jt-(T+1)}}$. In all panels of this figure g represents poor Whites, Blacks, Hispanics and Asians. We control for $\boldsymbol{v_{jt-T}}$ in all specifications. 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 $\boldsymbol{\theta}$ in any specification.

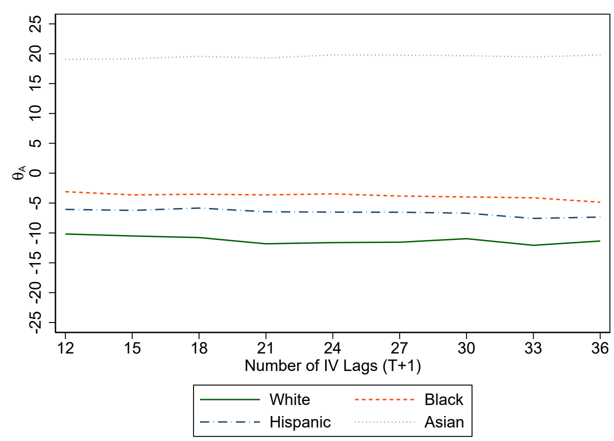
Figure 11: Responses of Poor Households of Different Races to Race and Income Compositions ($\theta_{g,g'}$) for Different Values of T: Price Explicitly Included



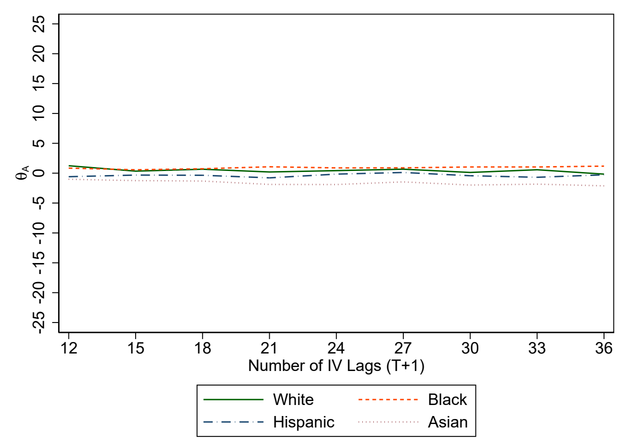
(a) Responses to Black share



(b) Responses to Hispanic Share



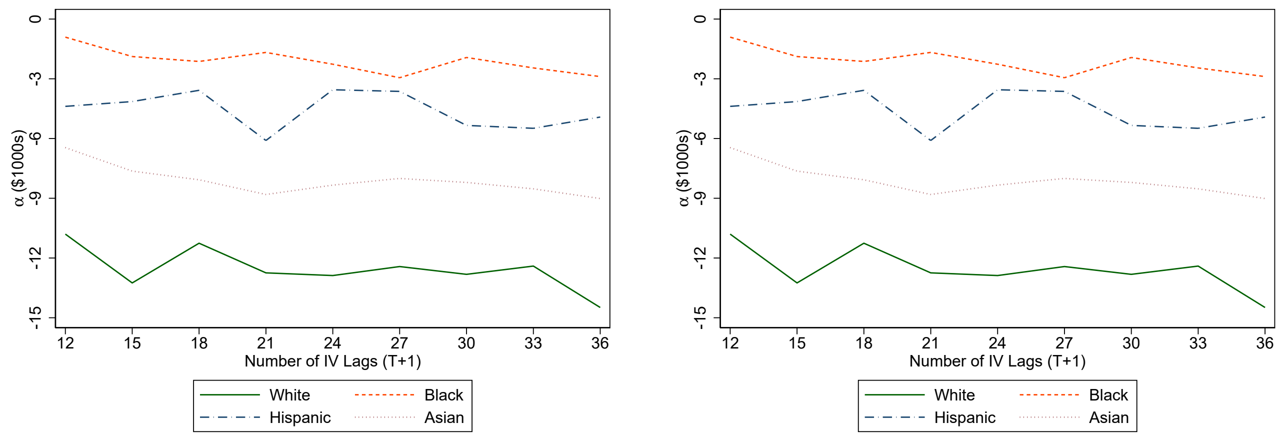
(c) Responses to Asian Share



(d) Responses to Poor Share

Notes: Each panel shows $\hat{\theta}_{g,g'}$ for all g' for different values of $T+1$, the lag of the Instrumental Variable, $\boldsymbol{s_{jt-(T+1)}}$. In all panels of this figure g represents poor Whites, Blacks, Hispanics and Asians. We control for $\boldsymbol{v_{jt-T}}$ in all specifications. 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 $\boldsymbol{\theta}$ in any specification.

Figure 12: Responses of Households to Prices (α_g) for Different Values of T: Price Explicitly Included



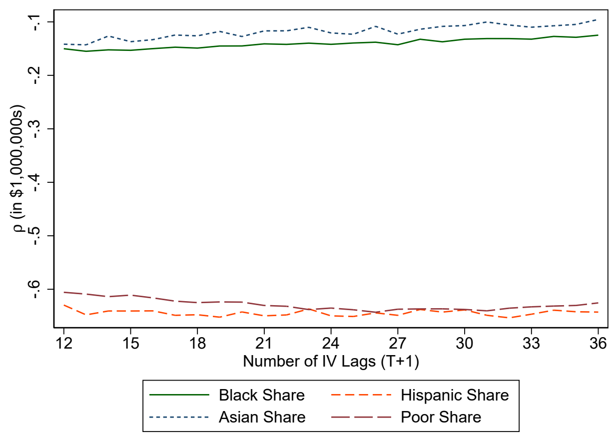
(a) Rich Households (b) Poor Households

Notes: The plot shows $\hat{\alpha}_g$ for all g for different values of T, the lag of the Instrumental Variables, $\boldsymbol{s_{jt-T}}$ and P_{jt-T}. 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 $\boldsymbol{\alpha}$ for all values of T.

B Incorporating Additional Information on Prices and Amenities

In equation (6) we intentionally specified only one type of observed amenity, the vector of socioeconomic compositions $\boldsymbol{s_{jt}}$. This is the natural variable of interest that a researcher would specify when studying socioeconomic segregation. Nevertheless, it is valuable to consider the case in which a researcher might observe other amenities as well. In this section, we describe additional empirical exercises that one can perform when such observables are available, and the conditions under which such exercises are feasible. A key takeaway is that additional observables do not allow us to relax the identifying assumptions that yielded our previous results. In fact, incorporating additional observables actually requires us to impose *further* identifying assumptions, although this does yield further results that illuminate our understanding of segregation.

Figure 13: Implicit Price of Race and Income Compositions ($\rho_{g'}$) for Different Values of T



Notes: Each panel shows $\hat{\rho}_{g'}$ for all g' for different values of $T+1$, the lag of the Instrumental Variable, $\boldsymbol{s_{jt-(T+1)}}$. In all panels of this figure g represents poor Whites, Blacks, Hispanics and Asians. We control for $\boldsymbol{v_{jt-T}}$ in all specifications. 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 $\boldsymbol{\rho}$ in any specification.

In our application, we observe one additional amenity: the average price of neighborhood home sales, P_{jt}.[33] To keep the discussion concrete, we extend our general framework to incorporate data on neighborhood prices, although the extension to any other amenity is analogous.

B.1 Empirical Framework with Price Data

B.1.1 Set Up

We modify our specification of expected utility in equation (5) as

$$v_{gjt}^{e} = \boldsymbol{\theta_g'} \boldsymbol{s_{jt}^{e}} + \alpha_g P_{jt}^{e} + \xi_{gjt}^{e,P} \tag{18}$$

where $P_{jt}^{e} = \mathbb{E}\left[P_{jt} | \boldsymbol{x_{gt}}\right]$ is the expected neighborhood price, and $\xi_{gjt}^{e,P}$ now excludes the expected valuation of the neighborhood due to price. The parameter vector $\boldsymbol{\theta_g}$ corresponds to group g responses to the expected socioeconomic composition of the neighborhood *net of price effects*, in contrast with $\boldsymbol{\beta_g}$, which corresponded to those responses inclusive of price effects.

[33]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).

Figure 14: Causal Graph of Socioeconomic Composition, Price and Valuation

$$\begin{tikzcd}
\boldsymbol{s^{e}_{g'jt}} \arrow[d, "\rho_{g'}"] \arrow[dr, "\theta_{g,g'}"] & \\
P^{e}_{jt} \arrow[r, "\alpha_g"] & v^{e}_{gjt}
\end{tikzcd}$$

Because the socioeconomic composition might be expected to be capitalized in prices, we need to include an additional equation:[34]

$$P^e_{jt} = \rho_0 + \boldsymbol{\rho}'\boldsymbol{s^e_{jt}} + \eta^e_{jt}. \tag{19}$$

The following figure illustrates how these two equations are related to each other. We represent causal relationships as directed arrows, and for simplicity we focus on the causal effects of a single element of the vector $\boldsymbol{s^e_{jt}}$, the expected share of group g' $s^e_{g'jt}$.

Group g households' valuation of neighborhood j are affected by their expectations of the group g' share in that neighborhood both directly (encapsulated by $\theta_{g,g'}$) and indirectly through the expected price response (encapsulated by $\rho_{g'}$ and α_g). Our baseline model without price in equation (5) incorporates this entire causal effect into the single parameter $\beta_{g,g'}$. As such, we can represent the mathematical relationship between the parameters of the model with price data and the parameters of the baseline model as

$$\beta_{g,g'} = \theta_{g,g'} + \rho_{g'}\alpha_g \tag{20}$$

For intuition, it is useful to explicitly describe a hypothetical sequence of events that might occur within a single period t. Consider a change in $s^e_{g'jt}$ at the very beginning of period t. As households of group g change their expectation regarding the share of group g', they will anticipate their value for the neighborhood (v^e_{gjt}) will change in two ways: first, directly depending on whether they value $s_{g'jt}$ per se, which is reflected in $\theta_{g,g'}$. Second, indirectly via prices. Indeed, households will anticipate changes in the

[34]Technically, it is possible that the effect of $\boldsymbol{s^e_{jt}}$ on P^e_{jt} changes also with g, since P^e_{jt} is the expectation of group g with respect to P_{jt}. Because we do not have data on price expectations, we cannot explicitly consider this possibility. If $\boldsymbol{\rho_g}$ represents this potentially heterogeneous effect, then η^e_{jt} in equation (19) includes the term $(\boldsymbol{\rho_g} - \boldsymbol{\rho})'\boldsymbol{s^e_{jt}}$.

neighborhood price due to the original change in $s^e_{g'jt}$ (via $\rho_{g'}$), which in turn should affect v^e_{gjt} via α_g. For each g, the coefficient $\beta_{g,g'}$ incorporates the effect of $\boldsymbol{s^e_{g'jt}}$ on v^e_{gjt} via these two channels. Note that so far, we make no assumptions on the particulars of the process by which $s^e_{g'jt}$ causes v^e_{gjt}. We simply posit that however households expect this process to be, it can be decomposed into one price channel, $\rho_{g'} \cdot \alpha_g$ and a second residual (i.e., every other channel beyond prices) channel $\theta_{g,g'}$. In particular, the mechanism that leads changes in their expected share of group g' ($s^e_{g'jt}$) to affect changes in expected prices (P^e_{jt}) is not specified; it may include expected adjustments in the supply-side or the demand-side, or both, and it may or may not include the expectation that the market will clear (i.e., actual demand equals actual supply).[35]

We re-write equation (18) from the perspective of the econometrician as

$$v^e_{gjt} = \boldsymbol{\theta'_g s_{jt}} + \alpha_g P_{jt} + \xi^P_{gjt} \tag{21}$$

where $\xi^P_{gjt} = \xi^{e,P}_{gjt} + \boldsymbol{\theta'_g}\left(\boldsymbol{s^e_{jt}} - \boldsymbol{s_{jt}}\right) + \alpha_g\left(P^e_{jt} - P_{jt}\right)$. Analogously, we re-write equation (19) from the perspective of the econometrician as

$$P_{jt} = \rho_0 + \boldsymbol{\rho' s_{jt}} + \eta_{jt}, \tag{22}$$

where $\eta_{jt} = \eta^e_{jt} + \boldsymbol{\rho'}\left(\boldsymbol{s^e_{jt}} - \boldsymbol{s_{jt}}\right)$.

B.1.2 Identification

Instead of Assumption 3, we make the following assumption to identify both $\boldsymbol{\theta_g}$ and α_g for all g.

Assumption 3′. *IV Validity (ξ). For all g and t, there exists some $T > 0$ such that*

$$\mathbb{E}\left[\xi^P_{gjt} \,\middle|\, \boldsymbol{v^e_{jt-T}}, \boldsymbol{s_{jt-(T+1)}}, P_{jt-(T+1)}\right] = \mathbb{E}\left[\xi^P_{gjt} \,\middle|\, \boldsymbol{v^e_{jt-T}}\right] \tag{23}$$

Assumption **3′** replaces Assumption 3 and states that the components of $s_{jt-(T+1)}$ and $P_{jt-(T+1)}$ that are orthogonal to $\boldsymbol{v^e_{jt-T}}$ are valid IVs for $\boldsymbol{s_{jt}}$ and P_{jt}. The argument for the plausibility of this validity assumption is analogous to the argument discussed in Section 3: information from the distant past ($\tau < t - T$) that turned out to be

[35]This hypothetical sequence of events is related to the sequence in Bayer, McMillan and Rueben (2004*a*) describing how steady state is assumed to be achieved in each period in their simulation. The key departure here is that we do not assume that these expectations are connected to what is observed in the data, so data may be out of steady state at any given period t.

irrelevant for the sorting decisions of the more recent past $(t-T)$ should not be relevant for sorting decisions in t. The argument for the relevance assumption of these IV is also analogous. In particular, note that P_{jt} is in part determined by households' willingness to *sell* their homes, which in turn depends on how mismatched to their neighborhood these households currently are. Since some of these mismatched households made their original sorting decision before $t-T$, P_{jt} should be correlated to information from the far past that turned out to be irrelevant more recently. For instance, the "unrealized" component of the price at which a household purchased their home in $\tau < t-T$ (i.e., the component due to $\boldsymbol{\omega_{gj\tau}} = \boldsymbol{x_{g\tau}} - \mathbb{E}\left[\boldsymbol{x_{g\tau}} \middle| \boldsymbol{v^e_{jt-T}}\right]$, see Section 3) may still influence the household's willingness to sell in t.

In order to identify $\boldsymbol{\rho}$ in equation (22), we make the following analogous assumption, which allows us to use the component of $\boldsymbol{s_{jt-(T+1)}}$ that is orthogonal to $\boldsymbol{v^e_{jt-T}}$ as an IV for $\boldsymbol{s_{jt}}$:

Assumption 6′. *IV Validity (η). For all t, there exists some $T > 0$ such that*

$$\mathbb{E}\left[\eta_{jt} \middle| \boldsymbol{v^e_{jt-T}}, \boldsymbol{s_{jt-(T+1)}}\right] = \mathbb{E}\left[\eta_{jt} \middle| \boldsymbol{v^e_{jt-T}}\right]$$

Assumption **6′** is closely related to Assumption **3′**. Note that η_{jt} incorporates the price capitalization of "demand-side" unobservables (ξ^P_{gjt} in equation (21)) as well as supply-side ones (not explicitly modeled in this paper, but potentially included in η_{jt}). Hence, Assumption **6′** states that the price capitalization of other amenities in t does not rely on the information from the distant past ($\tau < t-T$) that was ignored by all households who sorted in the more recent past $(t-T)$. The relevance of $\boldsymbol{s_{jt-(T+1)}}$ as an IV for $\boldsymbol{s_{jt}}$ conditional on $\boldsymbol{v^e_{jt-T}}$ follows analogously as well, i.e. due to mismatched households.

B.2 Estimation

Stage 1: Estimation of v^e_{gjt} and ϕ_g

This stage follows exactly as in the baseline case without price data presented in Section 3.2.

Stage 2: Estimation of $\boldsymbol{\theta_g}$, α_g and ρ_g

We re-write equation (21) based on observed quantities as

$$\hat{v}^e_{gjt} = \boldsymbol{\theta_g'} \boldsymbol{s_{jt}} + \alpha_g P_{jt} + \boldsymbol{\gamma_g'} \boldsymbol{\hat{v}^e_{jt-T}} + \underbrace{\xi^P_{gjt} + \hat{v}^e_{gjt} - v^e_{gjt} - \boldsymbol{\gamma_g'} \boldsymbol{\hat{v}^e_{jt-T}}}_{\text{error}_{gjt}}, \tag{24}$$

We estimate $\boldsymbol{\theta_g}$ in the equation above via Two Stage Least Squares using $\boldsymbol{s_{jt-(T+1)}}$ and $P_{jt-(T+1)}$ as an IV for $\boldsymbol{s_{jt}}$, controlling for $\boldsymbol{\hat{v}^e_{jt-T}}$. Based on Assumption 1, $\hat{v}^e_{gjt}$ converges to v^e_{gjt}, and $\boldsymbol{\hat{v}^e_{jt-T}}$ converges to $\boldsymbol{v^e_{jt-T}}$, and Assumption **3**$'$ implies $\boldsymbol{s_{jt-(T+1)}}$ and $P_{jt-(T+1)}$ are both uncorrelated to ξ^P_{gjt}, conditional on $\boldsymbol{v^e_{jt-T}}$.

Similarly, re-writing equation (22) to include control variable $\boldsymbol{v^e_{jt-T}}$, we obtain

$$P_{jt} = \rho_0 + \boldsymbol{\rho}' \boldsymbol{s_{jt}} + \boldsymbol{\lambda}' \boldsymbol{\hat{v}^e_{jt-T}} + \underbrace{\eta_{jt} - \boldsymbol{\lambda}' \boldsymbol{\hat{v}^e_{jt-T}}}_{\text{error}_{jt}}. \tag{25}$$

We estimate $\boldsymbol{\rho}$ in equation (22) via Two Stage Least Squares using $\boldsymbol{s_{jt-(T+1)}}$ as an IV for $\boldsymbol{s_{jt}}$, controlling for $\boldsymbol{\hat{v}^e_{jt-T}}$. Based on Assumption 1, $\boldsymbol{\hat{v}^e_{jt-T}}$ converges to $\boldsymbol{v^e_{jt-T}}$, and Assumption **6**$'$ implies $\boldsymbol{s_{jt-(T+1)}}$ is uncorrelated to η_{jt} conditional on $\boldsymbol{v^e_{jt-T}}$.

B.3 Simulation

Given counterfactual values of the parameter vector $(\tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\rho}})$ and counterfactual initial conditions $(\tilde{\boldsymbol{s}}^{\boldsymbol{0}}, \tilde{\boldsymbol{\eta}}^{\boldsymbol{0}}, \tilde{\boldsymbol{N}}^{\boldsymbol{0}}, \tilde{\boldsymbol{\xi}}^{\boldsymbol{P,0}}, \tilde{\boldsymbol{\eta}}^{\boldsymbol{0}})$, we obtain $\boldsymbol{s^{\tau+1}} = \boldsymbol{s}(\tilde{\boldsymbol{s}}^{\boldsymbol{\tau}})$ for each $\tau \geq 0$ using equations (15) and (16), as in the simulation without prices. However, instead of using equation (17), we update $v^e_{gj}(\tilde{\boldsymbol{s}}^{\boldsymbol{\tau}})$ according to

$$v^e_{gj}\left(\tilde{\boldsymbol{s}}^{\boldsymbol{\tau}}\right) = \tilde{\boldsymbol{\theta}}_{\boldsymbol{g}}' \tilde{\boldsymbol{s}}^{\boldsymbol{\tau}}_{\boldsymbol{j}} + \tilde{\alpha}_g P_j\left(\tilde{\boldsymbol{s}}^{\boldsymbol{\tau}}\right) + \tilde{\xi}^{P,0}{}_{gj}, \tag{26}$$

where

$$P_j\left(\tilde{\boldsymbol{s}}^{\boldsymbol{\tau}}\right) = \tilde{\boldsymbol{\rho}}' \tilde{\boldsymbol{s}}^{\boldsymbol{\tau}}_{\boldsymbol{j}} + \tilde{\eta}^0_j. \tag{27}$$

B.4 Interpretation of Model Parameters

Interpretation of $\boldsymbol{\theta}$ Although $\boldsymbol{\theta}$ is analogous to $\boldsymbol{\beta}$ in that it captures the various responses of households of different groups to their expectations of the socioeconomic compositions of neighborhoods, it has a subtly different interpretation. $\beta_{g,g'}$ represents the total marginal effect of an increase in the expected g' share on the group g valuation of a neighborhood. This effect includes both taste-based and statistical discrimination.

In contrast, $\theta_{g,g'}$ represents the total marginal effect of an increase in the expected g' share on the group g valuation of a neighborhood *conditional on expected prices*. That is, this effect excludes the component of statistical discrimination due to any expected change in price that would be induced by the increase in g' share. This distinction is particularly important when considering "no discrimination" counterfactuals where certain elements of $\boldsymbol{\theta}$ are set to zero. Because $\boldsymbol{\rho}$ and $\boldsymbol{\alpha}$ are not set to zero in these counterfactuals, group g valuations may endogenously change in response to an increase in expected g' share through the price channel. Hence, these should be understood as "no discrimination except for statistical discrimination due to the expected effects of socioeconomic composition on prices" counterfactuals.

Interpretation of α As in the case of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, the parameter $\boldsymbol{\alpha}$ incorporates two components: $\alpha_g = \frac{\partial v^e_{gjt}}{\partial P_{jt}} = \frac{\partial u^e_{gjt}}{\partial P_{jt}} + \frac{\partial CV^e_{gjt}}{\partial P_{jt}}$. The flow utility component should be negative since households prefer to pay lower prices for their house, all else constant. However, the sign of the continuation value component is theoretically ambiguous since it depends on whether the price of a neighborhood today signals disproportionate expected future appreciation relative to an otherwise comparable neighborhood, which would have consequences for households' expected wealth.[36]

Interpretation of $\boldsymbol{\rho}$ The price response parameter, $\boldsymbol{\rho}$, incorporates both supply-side and demand-side considerations, though we are agnostic as to their particulars. $\boldsymbol{\rho}$ should simply be understood as the best linear approximation of the (potentially non-linear) causal expectation function (CEF) of $\boldsymbol{s_{jt}}$ on P_{jt} (Angrist and Pischke (2009)). We do not invoke the assumptions that allow this equation to be interpreted as a hedonic price equation from which tastes or costs can be recovered (Rosen (1974)), nor do we invoke the assumptions that allow $\boldsymbol{\rho}$ to be interpreted as a marginal willingness to pay parameter. For instance, $\boldsymbol{\rho}$ would equal zero if household demand was unresponsive to changes in $\boldsymbol{s_{jt}}$, or if the supply of housing was perfectly elastic. We would be unable to distinguish between these two scenarios given an estimate of $\hat{\boldsymbol{\rho}} = \mathbf{0}$, but this would

[36] The buying and selling of a house may impact household wealth. Despite its undeniable importance when studying the behavior of households, we do not explicitly model the effects of moving on wealth, and we do not allow for household heterogeneity by wealth either. 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 households of each race and income level to study their decisions by wealth levels. Note, however, that wealth is partially incorporated in our analysis since our parameters are allowed to vary by group, and these groups may have different wealth on average.

not affect our counterfactuals.

Interpretation of $\boldsymbol{\eta}$ The unobserved terms $\boldsymbol{\eta}$ in equation (22) encompass every determinant of price other than the endogenous factor $\boldsymbol{s}$. Thus, by keeping $\tilde{\boldsymbol{\eta}}_{\boldsymbol{jt}}$ fixed at $\hat{\boldsymbol{\eta}}^{\boldsymbol{P}}_{\boldsymbol{j\mathcal{T}}}$ when considering counterfactuals in the model with price, we allow P_{jt} to change only insofar as it evolves endogenously via $\boldsymbol{s}$.

Interpretation of $\boldsymbol{\xi^P}$ The unobserved terms $\boldsymbol{\xi^P}$ in equation (21) and $\boldsymbol{\xi}$ in equation (6) differ in the following way: $\boldsymbol{\xi^P}$ does not incorporate the component of $\boldsymbol{\xi}$ that encompasses sorting purely on the basis of exogenous price changes (i.e., due to $\boldsymbol{\eta}$). However, *both* terms exclude the endogenous component due to sorting on the basis of $\boldsymbol{s}$ that is mediated via prices (see Figure 14). As explained above, this component is incorporated instead in $\beta_{g,g'}$ or $\rho_{g'} \cdot \alpha_g$, respectively. Thus, when considering counterfactuals in the model with prices, by keeping $\tilde{\boldsymbol{\xi}}^{\boldsymbol{P}}_{\boldsymbol{gjt}}$ fixed at $\hat{\boldsymbol{\xi}}^{\boldsymbol{P}}_{\boldsymbol{gj\mathcal{T}}}$, we allow v^e_{gjt} to change only insofar as it evolves endogenously via $\boldsymbol{s}$ directly or indirectly (through P).

B.5 Estimation Results with Price Data

We present estimates of $\boldsymbol{\theta}$, $\boldsymbol{\pi}$, $\boldsymbol{\rho}$ and $\boldsymbol{\phi}$ in Table 4. The signs and relative magnitudes of our estimates of $\boldsymbol{\theta}$ are roughly similar to the estimates of $\boldsymbol{\beta}$ from our baseline model with the exception that rich households of all races respond less intensely to poor neighbors than before. This suggests that much of the response to poor households that was identified in our main analysis reflected differential responses to neighborhood prices as opposed to the income of neighbors. All households respond negatively to higher neighborhood prices, but poor White and Asian households are over eight and three times more price-sensitive than their rich counterparts respectively. Our estimates of $\boldsymbol{\rho}$ presented in the far right column 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 \$12,300. This effect is roughly the same size as the effect of an increase in the expected Asian share of a neighborhood and about four times smaller than the effects of increases in the expected Hispanic and poor shares. In addition, we cannot reject that $\beta_{g,g'} = \theta_{g,g'} + \boldsymbol{\rho'_{g'}\alpha_g}$ for all g, g', which is consistent with the validity of identifying Assumptions 3, $\mathbf{3'}$ and $\mathbf{6'}$.

Table 4: Responses to the Socioeconomic Compositions and Prices of Neighborhoods

	White		Black		Hispanic		Asian		
	Rich	Poor	Rich	Poor	Rich	Poor	Rich	Poor	Prices
Responses to:									
Black Share	-9.82	-9.17	8.89	10.17	-1.84	2.84	-3.67	-2.27	-0.123
	(0.38)	(0.47)	(0.41)	(0.40)	(0.35)	(0.35)	(0.36)	(0.36)	(0.010)
Hispanic Share	-16.08	-13.78	-1.97	-1.85	22.05	25.05	2.13	0.67	-0.563
	(0.55)	(0.72)	(0.53)	(0.52)	(0.61)	(0.61)	(0.64)	(0.64)	(0.012)
Asian Share	-4.49	-11.66	-1.34	-2.80	-2.42	-4.60	18.24	20.64	-0.094
	(0.34)	(0.42)	(0.36)	(0.33)	(0.44)	(0.44)	(0.55)	(0.55)	(0.013)
Poor Share	-5.81	-3.44	-7.24	0.55	-12.69	-3.02	-13.45	-4.49	-0.584
	(0.47)	(0.50)	(0.36)	(0.32)	(0.39)	(0.39)	(0.45)	(0.45)	(0.013)
Response to Price	-1.36	-12.72	-3.45	-1.39	-6.45	-5.03	-2.04	-7.12	
(Thousands)	(0.38)	(0.90)	(0.64)	(0.46)	(0.07)	(0.08)	(0.70)	(0.74)	
Moving Costs	28.57	28.70	27.44	27.60	28.04	28.16	28.06	27.64	
	(0.01)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)	
Num. of Observations					147,840				36,960

Notes: π_g is allowed to vary only by income group. We use $s_{g'jt-13}$ for all g' and P_{jt-13} as instrumental variables. White is the omitted racial share and rich is the omitted income share. Price as a dependent variable is denominated in millions of dollars. Moving costs are estimated by GMM (see equation 13). All standard errors clustered by group-month. The p-values for both the Cragg-Donald and the Kleinbergen-Paap weak identification tests for both regressions are less than 0.001, which imply strong first stages.

Sensitivity Analysis

In Appendix A, we report estimates of $\boldsymbol{\theta}$, $\boldsymbol{\rho}$ and $\boldsymbol{\alpha}$ for different values of $T+1$ (the period corresponding to our IVs) in Figures 10 through 13. Larger values of $T+1$ relax the validity assumptions. We find that estimates of every element of $\boldsymbol{\theta}$, $\boldsymbol{\rho}$ and $\boldsymbol{\alpha}$ changes little for $T+1=13,\ldots,36$. This evidence, coupled with the evidence that $\beta_{g,g'}=\theta_{g,g'}+\rho_{g'}\alpha_g$ is inferred to be valid for all g and g', even though $\hat{\boldsymbol{\beta}}$ is obtained under weaker assumptions than $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\rho}}$ and $\hat{\boldsymbol{\alpha}}$, is consistent with the idea that the IVs are valid in our setting.

B.6 Simulation Results with Price Data

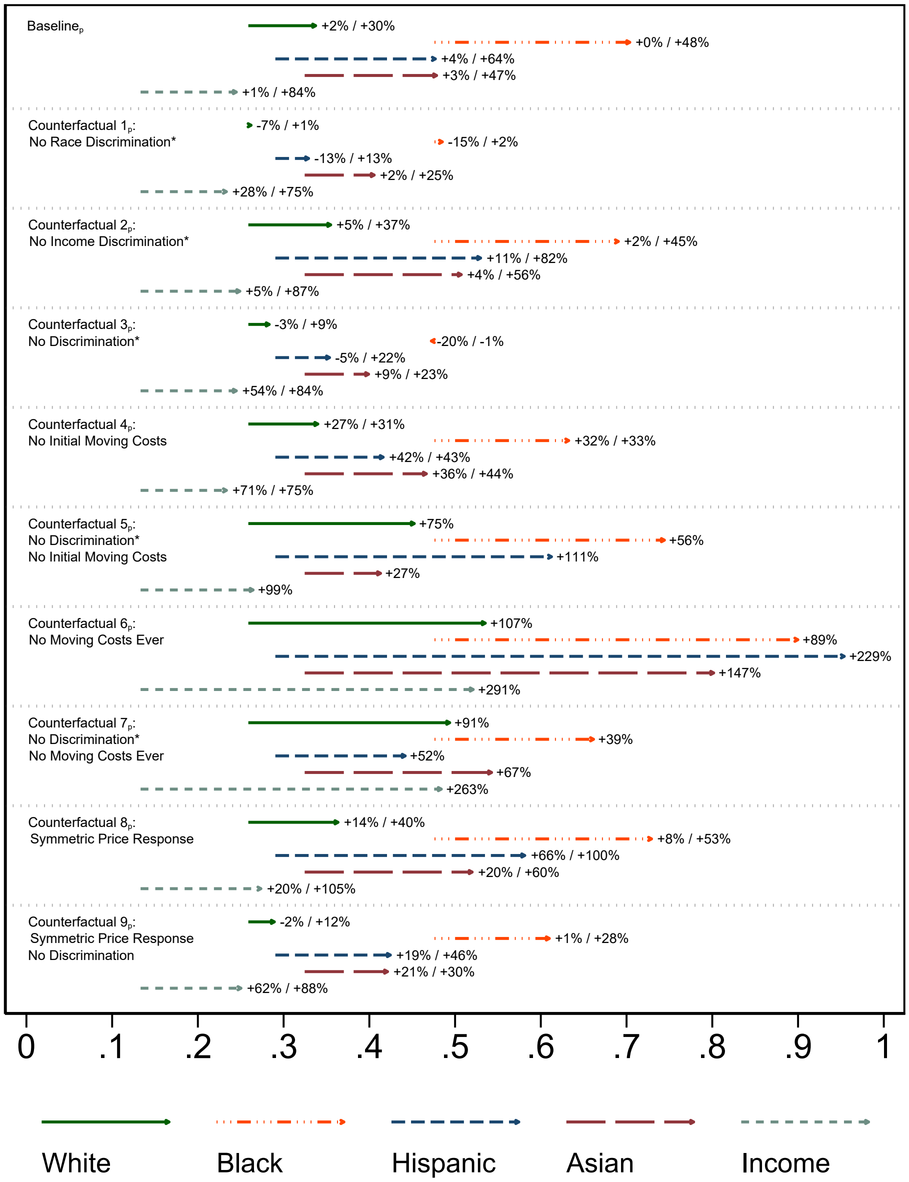
In Figure 15, we present the simulated change in segregation levels under all previously considered counterfactuals plus two new counterfactuals. A subscript of p indicates that the simulation is performed using our model with price. As expected, our baseline results from models with and without price data are effectively unchanged since we cannot reject $\beta_{g,g'}=\theta_{g,g'}+\boldsymbol{\rho}'_{\boldsymbol{g}'}\boldsymbol{\alpha}_{\boldsymbol{g}}$ for all g,g'. We conclude that explicitly adding price to the model does not add any predictive power to explain the baseline trend in segregation.

Next, we consider counterfactuals where households are race-blinded (1_p), income-blinded (2_p), or both (3_p). We note that these counterfactuals are different from their analogs 1, 2, and 3 in Figure 3, since $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are conceptually different parameters. While in counterfactuals 1-3 we set elements of $\tilde{\boldsymbol{\beta}}$ to zero, in counterfactuals 1_p-3_p we set elements of $\tilde{\boldsymbol{\theta}}$ to zero instead. Thus, in counterfactuals 1_p-3_p there is still room for discriminatory sorting on the basis of prices via $\boldsymbol{\rho}'_{\boldsymbol{g}'}\boldsymbol{\alpha}_{\boldsymbol{g}}$. A comparison among these counterfactuals reveal that the price channel does not interact much with racial discrimination, but it does interact with income discrimination. This is expected, since the elements of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$ with regard to race are more similar than the elements of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$ with regard to income (Tables 2 and 4). In particular, a comparison between counterfactuals 2_p and 2 (as well as 3_p and 3) leads us to conclude that about half of the long-run increase in income segregation is attributable to the component of income discrimination due to prices.

When we eliminate initial moving costs in counterfactuals 4_p and 4 we obtain nearly identical results, and the same is true when we eliminate moving costs in perpetuity in counterfactuals 6_p and 6. This follows from the fact that the baseline simulations are very similar, hence their full discriminatory mechanisms should have similar effects at any level of moving costs. However, when we compare counterfactuals 5_p and 5, we find some differences, leaving us to conclude that discrimination via prices ($\boldsymbol{\rho'_{g'}\alpha_g}$) would play some role absent moving costs. Specifically, without initial moving costs, income segregation would be smaller because of discrimination due to prices. Thus far, we have concluded that whenever discrimination reduces long-run segregation (instead of increasing it), it is because uncertainty plays an important role mediating the discriminatory sorting process. This case is no different: a comparison between counterfactuals 7_p and 7 shows that when we make uncertainty no longer an issue (by allowing household to costlessly move as many times as they want in order to resolve *ex post* mismatch that would have arisen due to uncertainty), discrimination due to prices actually increases income segregation, as expected.

As noted, the model with prices allows us to perform two additional counterfactuals. In counterfactual 8_p, we modify baseline$_p$ by equalizing all price response parameters ($\tilde{\alpha}_g = \tilde{\alpha}$, where $\tilde{\alpha}$ is equal to a population weighted sum of $\hat{\alpha}_{\mathrm{g}}$ across all g) and find that segregation would increase much more quickly and to higher long-run levels. In counterfactual 9_p, we modify counterfactual 3_p in the same way and obtain analogous results. These two results suggest that differential price responses mitigate discriminatory sorting because neighborhood characteristics are bundled, i.e., sorting on the basis of prices tends to lead the household to a different neighborhood than other types of sorting (both discriminatory and non-discriminatory). For example, group g households may find neighborhood j disproportionately more attractive than j' (relative to group g' households) because of ξ^P, and they may find neighborhood j' disproportionately more attractive than j because of P. It is also useful to compare counterfactuals 8_p and 9_p. This allows us to conclude that shutting off discriminatory channels (excluding price) leads to lower medium- and long-run levels of segregation, as expected.

Figure 15: Medium- and Long-Run Changes in Segregation (Simulated)



Notes: The arrows represent the changes in simulated dissimilarity indices for households of each race/income from November 2004 onward in the absence of shocks to ξ. (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 numbers shown correspond to medium-run (5 years into the future) and long-run (after neighborhoods have reached steady state) changes in segregation relative to November 2004. Medium-run and long-run changes are identical in counterfactuals 5_p-7_p. Details of each counterfactual are presented in Section 4.2.

The effects of endogenous price responses on segregation dynamics are presented in Figure 16. A comparison of Figure 16 with Figure 2 and the third panel of Figure 4 shows that symmetric price responses lead to faster convergence to steady state with and without discrimination. This once again reflects the fact that neighborhood characteristics are bundled. Heterogeneous price responses are a countervailing force to sorting on the basis of other amenities in $\boldsymbol{\xi}^P$: households may find neighborhood j more attractive than j' because of $\boldsymbol{\xi^P}$ and neighborhood j' more attractive than j because of P, which could prolong convergence to steady state.

Finally, the model with prices allows us to consider counterfactual trajectories of prices in our simulations. We present these results in Figure 17. As shown in the baseline counterfactual, the socioeconomic sorting of households across neighborhoods has a negligible impact on prices in the medium-run. In the long-run, socioeconomic sorting does cause more substantial price appreciation in the most expensive neighborhoods, but only on the order of 15% over the course of a century and a half. This pales in comparison with the large price appreciation that we observe in the Bay Area, which suggests that other amenities loaded in $\boldsymbol{\xi^P}$ are more relevant by far to explain housing price dynamics. When moving costs are eliminated, we unsurprisingly find greater dispersion in house prices, as sorting is less restricted.

C Monte Carlo Simulations

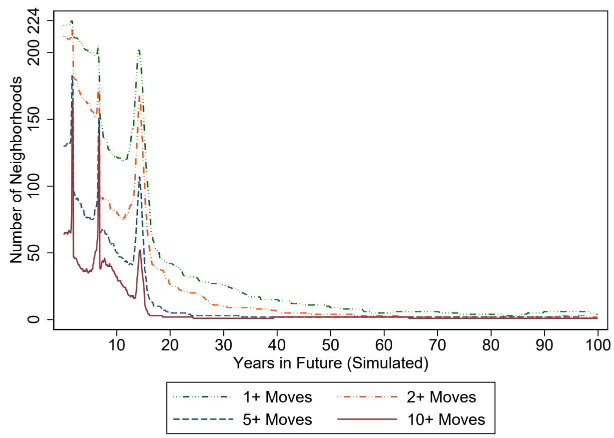
C.1 Set up

We consider the exact model from Section 3 with $G = 2$ instead of $G = 8$ for simplicity and clarity. We denote the two groups as A and B. For each sample, $N_{A0} = 1,000$, $N_{B0} = 9,000$, $J = 224$ and $\bar{t} = 179$ (the values of J and $\bar{t}$ correspond to the values in our data). We also chose parameter values for the data generation process that led the serial correlation of both v^e_{gjt} and s_{jt} to approximately match the serial correlation observed in the data in short-run ($cov(v^e_{gjt}, v^e_{gjt-1})$), medium-run ($cov(v^e_{gjt}, v^e_{gjt-60})$) and long-run ($cov(v^e_{gjt}, v^e_{gjt-279})$). Each sample is created as follows:

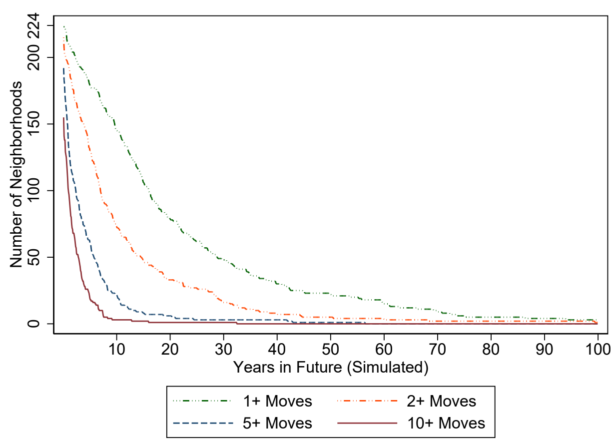
1. Draw ξ_{gj0} from $\xi_{gj0} \sim \mathcal{N}(0, 4)$ for each g and $j \geq 1$, and normalize $\xi_{g00} = 0$.

Figure 16: Number of Neighborhoods In Flux - Alternative Price Responses (Simulated)

(a) Symmetric Price Responses

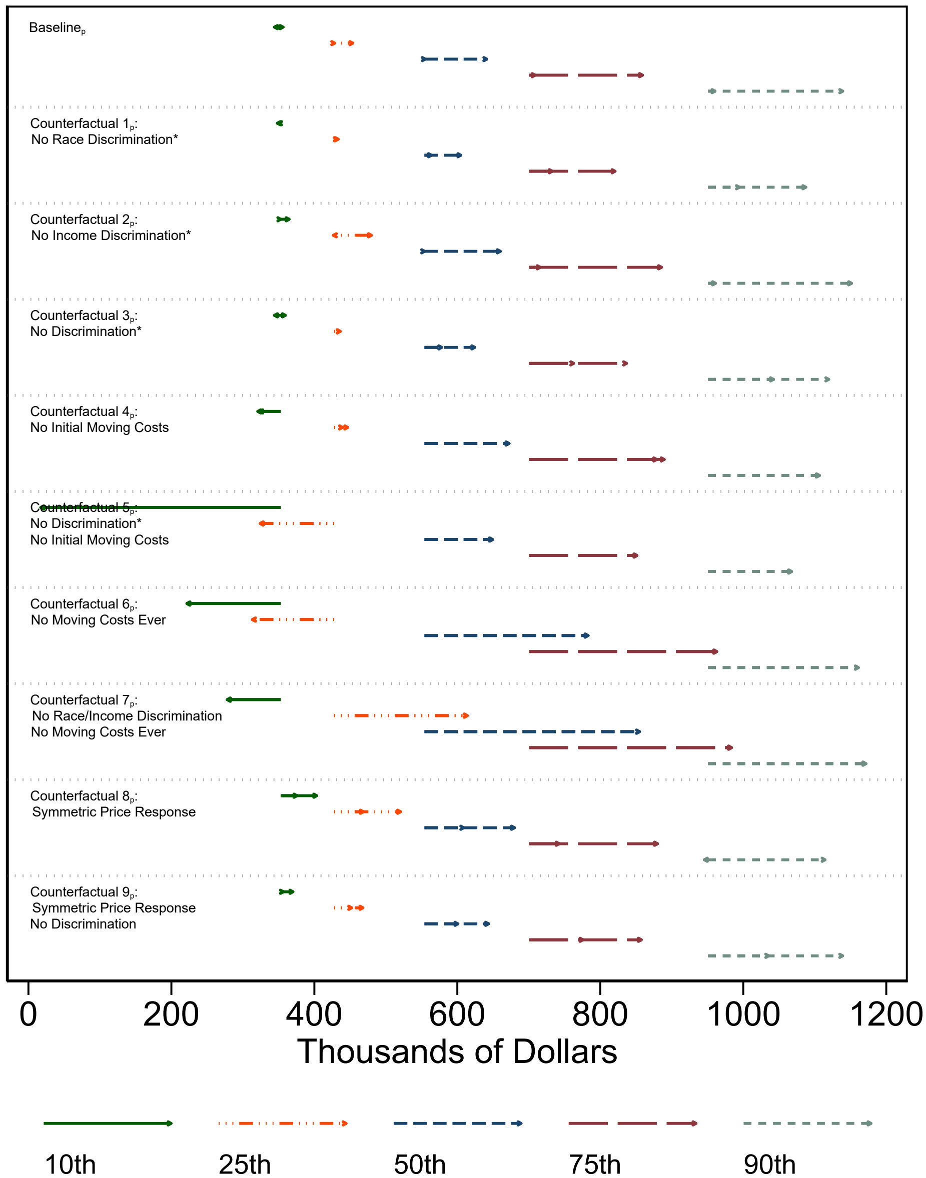


(b) Symmetric Price Responses, No Racial or Income Discrimination



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 17: Medium- and Long-Run Changes in Average Neighborhood Prices



Notes: The arrows represent the changes in simulated average neighborhood prices at various quantiles from November 2004 in the absence of exogenous shocks. The first arrowhead corresponds to a "medium-run" (5 years into the future), and the second arrow corresponds to a "long-run" (after neighborhoods have reached steady state). The medium-run and long-run values from counterfactuals 5_p-7_p are the same. All prices are in constant 2004 dollars. Details of each counterfactual are presented in Section B.

2. For $j \geq 1$, calculate $s_{j0} = \frac{N_{Aj0}}{N_{Aj0}+N_{Bj0}}$, where N_{gj0} is obtained by

$$N_{gj0} = N_{g0} \cdot \frac{\exp(\xi_{gj0})}{\sum_{k=0}^{J} \exp(\xi_{gk0})}.$$

   i.,e., households sort only on the basis of ξ in period 0. Steps 1 and 2 create the initial conditions, which are used to simulate the sample for $t \geq 1$. Data for $t = 0$ is not included as part of the sample when we perform estimation.

3. For $t \geq 1$ and $j \geq 1$, we draw ψ_{gjt} from $\psi_{gjt} \sim \mathcal{N}(0, 1)$.

4. For $t \geq 1$ and $j \geq 1$, obtain ξ_{gjt} from the following equation: $\xi_{gjt} = 0.8\xi_{gjt-1} + \psi_{gjt} - 0.8\psi_{gjt-1}$.

5. For $t \geq 1$, obtain $s_{jt} = \frac{N_{Ajt}}{N_{Ajt}+N_{Bjt}}$, where N_{gjt} is obtained by

$$\begin{aligned} N_{gjt} = \quad & N_{gjt-1} \times \left(\frac{\exp\left(v^e_{gjt}\right)}{\sum_{j'=1}^{J} \exp\left(v^e_{gj't} - \phi_g\right) + \exp\left(v^e_{gjt}\right)} \right) + \\ & + \sum_{k=1}^{J} N_{gkt-1} \times \left(\frac{\exp\left(v^e_{gjt} - \phi_g\right)}{\sum_{j'=1}^{J} \exp\left(v^e_{gj't} - \phi_g\right) + \exp\left(v^e_{gkt}\right)} \right) \end{aligned}$$

   with $v^e_{g0t} = 0$ and $v^e_{gjt} = \beta_g \cdot s_{jt} + \xi_{gjt}$ for $j \geq 1$. We set $\phi_A = 20$, $\phi_B = 15$, $\beta_A = 3$ and $\beta_B = -3$.[37]

These steps create a sample of N_{gjt} for $g = A, B$, $j = 1, ..., J$, and $t = 1, ..., \bar{t}$. This is one sample. We repeat these steps to create a total of 1,000 samples.

C.2 Monte Carlo Results

Table 5 shows the Monte Carlo results. For each different estimator, we show both the bias and the standard deviation, which are the two inputs in the Mean Square Error calculation ($MSE = Bias^2 + SD^2$).

The first row shows the results for the moving cost estimator of ϕ_g. It performs well, with small biases, which is to be expected given that the Monte Carlo assumes the dynamic model from Section 3.1.

The remaining rows of the table show the results for the discriminatory response estimators of β_g. For context, we start with the naive OLS estimator, which is biased

[37] Allowing for changes over time in the total number of households of a given group across all J neighborhoods does not change the results, as expected, since such variation is absorbed by group-month fixed effects.

for both β_A and β_B, as expected. The other rows show the performance of the 2SLS estimator of β_g using $s_{jt-(T+1)}$ as an IV for s_{jt}, controlling for $\boldsymbol{v^e_{jt-T}}$. We consider different values of T to show how the results tend to be fairly stable as T grows. Irrespective of the lag, the IV estimators of β_A and β_B tend to perform well.

Table 5: Monte Carlo Results: Estimates of ϕ_g and β_g

	Group A		Group B	
	Bias	SD	Bias	SD
ϕ_g: $\phi_A = 20$, $\phi_B = 15$	0.0002	0.0001	0.0001	0.0002
β_g: $\beta_A = 3$, $\beta_B = -3$				
OLS: $v^e_{gjt} = s_{jt}\beta_g + \xi_{gjt}$	0.1900	0.0593	-0.2103	0.0630
IV: $s_{jt-(T+1)}$ as instrument for s_{jt} inv$^e_{gjt}$ = $s_{jt}\beta_g + \boldsymbol{\gamma}' \cdot \boldsymbol{v^e_{jt-T}} + \text{error}_{gjt}$				
s_{jt-13} as IV, $\boldsymbol{v^e_{jt-12}}$ as controls	-0.0039	0.0292	0.0045	0.0290
s_{jt-19} as IV, $\boldsymbol{v^e_{jt-18}}$ as controls	0.0040	0.0294	-0.0024	0.0285
s_{jt-25} as IV, $\boldsymbol{v^e_{jt-24}}$ as controls	0.0058	0.0297	-0.0046	0.0288
s_{jt-31} as IV, $\boldsymbol{v^e_{jt-30}}$ as controls	0.0072	0.0311	-0.0040	0.0311
s_{jt-37} as IV, $\boldsymbol{v^e_{jt-36}}$ as controls	0.0080	0.0319	-0.0043	0.0320

Notes: Each specification includes group-month fixed effects as controls. Group B is the omitted group share.