Why Are Minorities Disproportionately Concentrated in Low-Quality Nursing Homes?

Alden Cheng

October 9, 2023

Abstract

This paper studies the underlying mechanisms giving rise to high levels of racial segregation and disparities across nursing homes in the US. Descriptively, I find that while residential segregation is an important explanation for racial segregation across nursing homes, it struggles to explain disparities. I provide reduced form evidence for several other explanations for disparities: nursing homes seem to discriminate against minorities in their admission practices, individuals tend to choose nursing homes with a higher share of residents of their own race, and minorities seem less sensitive to a high-profile information intervention in the form of the introduction of the star ratings system. To disentangle and quantify the effects of these proposed mechanisms, I then estimate a structural model. Counterfactual simulations confirm that residential segregation is indeed the main explanation for high levels of segregation, yet it explains little of disparities. By contrast, information frictions is the main driving force behind disparities but is largely orthogonal to segregation. Discrimination by nursing homes and in-group preferences play smaller roles.

1 Introduction

Racial segregation is a pervasive phenomenon in a number of important settings, such as school (Billings, Demming and Rockoff 2014), neighborhood (Card, Mas, and Rothstein 2007), and nursing home choice (Mack et al. 2020). Moreover, in many of these cases, minorities tend to be disproportionately concentrated in lower-quality institutions, leading to concerns over racial disparities. Past studies have explored various explanations for these segregation patterns in isolation, including in-group preferences, discrimination, and access, but it is difficult to compare the relative importance of

I am very grateful to Josh Angrist, David Autor, John Bowblis, Viola Corradini, Amy Finkelstein, Jon Gruber, Ben Handel, Riley League, and Kosali Simon for their valuable comments.

these explanations. In addition, studies do not always make a clear distinction between segregation and racial disparities, even though policy prescriptions may differ depending on which of these two issues the policymaker prioritizes.

In this paper, I take a first step to filling these gaps in the research by studying segregation and racial disparities in the nursing home context, where more than half of elderly Americans are expected to be admitted sometime during their lives (Hurd, Michaud, and Rohwedder 2017). Using a rich administrative data set on the universe of nursing home residents, I start by establishing several pieces of descriptive evidence on the structure and potential causes of segregation and disparities. To disentangle these explanations, I then estimate a structural model and conduct counterfactual simulations to quantify each of their contributions.

I start by showing that while residential segregation is likely an important contributor to nursing home segregation, it cannot fully explain racial disparities, as measured by the gap in quality of nursing homes chosen by different racial groups. Most residents choose a nursing home relatively close to where they live, so high levels of residential segregation feeds into nursing home segregation: indeed, segregation measured at a local level that better approximates residents' choice sets is far lower than statewide measures of nursing home segregation. However, even conditional on zip code of prior residence, minorities tend to be admitted to lower-quality nursing homes, suggesting that residential segregation is not the entire story when it comes to racial disparities.

I then present evidence that several alternative explanations may also contribute to racial disparities. First, nursing homes seem to discriminate against minority residents in their admissions practices: they are less likely to admit minority residents during times of capacity strain (when the option value of an empty bed is higher).[1] If higher-quality nursing homes are in higher demand and experience greater capacity strain, such admissions practices may give rise to racial disparities.

Second, residents seem to prefer nursing homes with a higher share of residents of their own race: dynamic panel and event study estimates both show that a shock to the share of minority admissions to a nursing home has a persistent effect. Such in-group preferences are often discussed in relation to their effects on segregation (Schelling 1971; Card, Mas, and Rothstein 2007), but may also contribute to the perpetuation of racial disparities given an initial distribution of racial groups across nursing homes with minorities disproportionately living in lower-quality nursing homes. However, this does

[1] I do not take a stance on whether this behavior is due to taste-based or statistical discrimination of a profit-maximizing firm (although I do control for other resident characteristics that may be associated with both race and their profitability to nursing homes).

not provide a fully satisfactory explanation for racial disparities in the sense that it cannot rationalize how the initial distribution came about.

Third, I provide evidence that minorities may face greater information frictions about nursing home quality than whites do. In particular, I exploit the introduction of nursing home star ratings at the end of 2008, which can be viewed as an information intervention. Unequal access to information may arise from the fact that information on various measures of nursing home quality is available on the CMS website, but internet use among racial minorities is lower than among whites (Jain et al. 2021). Indeed, while demand for higher-quality nursing homes increased following the introduction of star ratings, difference-in-differences estimates show that "improvements in choice" tended to be greater for white residents.

A shortcoming of the reduced form evidence is that it struggles to disentangle these different explanations for disparities, since observed choices can either reflect residents' preferences or rejections by other nursing homes (which are not observed in the data). Therefore, in the second part of the paper I estimate a structural model based on data from Florida between 2008–2010,[2] and conduct counterfactual simulations to quantify the contribution of each explanation to segregation and disparities. Specifically, the model incorporates distance, racial preferences, and information frictions in residents' decision utility, and occupancy, race, and other resident characteristics in nursing homes' admissions rule. The structural estimates provide additional support for the reduced form evidence: discrimination, in-group preferences, and information frictions all seem to play some role.

In terms of magnitude, counterfactual simulations indicate that residential segregation is by far the most important explanation for statewide segregation, but that it plays a smaller role in explaining racial disparities. By contrast, information frictions is the main driver of racial disparities, but is largely orthogonal to segregation, while in-group preferences and discrimination by nursing homes play a smaller role.[3]

This paper is linked to a vast literature on racial segregation and disparities. A number of previous studies have produced credible evidence on various explanations for these patterns. For example, Card, Mas and Rothstein (2007) show support for one of the key implications of Schelling's model

[2]I limit the sample due to computational constraints, and chose to focus on Florida due to demographic reasons, in particular its relatively old population, and substantial fraction of hispanic residents (which allows me to study the black and hispanic minority groups separately).

[3]It is important to emphasize that I only focus on one specific form of discrimination. For example, the greater information frictions that minority residents face may be the result of systematic discrimination more broadly (e.g. education, and internet access). Similarly, while I measure disparities based on nursing home *choice*, provision of lower-quality care to minorities and selective discharge practices by nursing homes (Hackmann, Pohl, and Ziebarth 2020) may further contribute racial disparities in resident *outcomes*.

of in-group preferences — the existence of tipping points — in the context of neighborhood choice, while Derenoncourt (2022) highlights the role of discrimination and location in explaining patterns of segregation and disparities that result from broad migration patterns. In addition, Billing, Demmings, and Rockoff (2014) establish a direct link between segregation and outcomes disparities, using a natural experiment induced by the end of race-based busing in Charlotte-Mecklenburg schools (CMS).

My paper contributes to this literature by studying several of these explanations simultaneously within a single setting, since most previous studies have focused on one explanation in isolation. This allows me to compare their effects on segregation and disparities, as well as potential interaction effects between various counterfactual policies. Moreover, by studying multiple forces in concert, I show that policies which reduce segregation may be different than those that reduce racial disparities.[4,5]

More narrowly, this paper is related to a literature on racial segregation and disparities in healthcare settings (Baicker et al. 2004; Smith et al. 2007; Rahman and Foster 2015). With the exception of the last reference, most of these studies tend to be reduced form and descriptive. As I discuss in greater detail later in the paper, reduced form evidence often struggles to disentangle demand and supply side explanations, so in this study I estimate a structural model in order to uncover the underlying mechanisms. The closest study to this paper is Rahman and Foster, who also study the role of in-group preferences, location, and preference heterogeneity in the context of nursing homes. However, they do not model nursing homes' admissions decisions, so their structural estimates reflect both residents' preferences and nursing homes' admission decisions.[6]

This paper proceeds as follows. In section 2, I provide some background on nursing homes and introduce my main data sources, before laying out some descriptive statistics on racial segregation and disparities. In section 3, I present reduced form evidence for various explanations for segregation and disparities. In section 4, I estimate an empirical matching model that incorporates all of these elements, and in section 5, I conduct counterfactual simulations to quantify the importance of these

[4] Clearly, under perfect integration, the average quality of nursing homes chosen by majority and minority groups will be the same. However, a reduction in segregation that does not achieve full integration need not necessarily reduce the gap in quality chosen by majority and minority groups. To see why this is the case, consider a simplified example with 3 nursing homes, A, B, and C, with cardinal quality measures given by 5, 2, and 1 respectively. There are 200 residents from the majority group, and 100 residents from the minority group, and initially all minority group residents reside in nursing home B, whereas majority group residents are split equally between nursing homes A and C. Now, suppose that 50 minority residents move from nursing home B to C, and 50 majority residents move from nursing homes C to B. This achieves a reduction in segregation, but the average gap in the quality of nursing homes chosen by majority and minority groups increases from $(5+1)/2-2=1$ to $(5/2+2/4+1/4)-(2+1)/2=7/4$.

[5] It should be noted that there are some crucial differences between the nursing home setting and those that were studied previously, such as education. Perhaps most meaningfully, peer effects are less likely to play an important role in nursing home settings compared to education: the racial composition of residents in an individual's nursing home is unlikely to have a direct causal impact on her own health. Such differences should be kept in mind when generalizing the findings in this paper to other settings.

[6] In addition, this paper provides several pieces of reduced form evidence supporting the underlying mechanisms.

factors for explaining racial segregation and disparities. Section 6 concludes.

2 Background

There are roughly 15,000 nursing homes in the US providing care for about 1.3 million Americans (CDC), and an estimated 56 percent of Americans aged 57–61 are expected to spend at least one night in a nursing home during their lifetimes (Hurd, Michaud, and Rohwedder 2017). However, a large literature has documented substantial segregation across nursing homes (see the meta-analysis by Mack et al. 2020), and that minorities tend to disproportionately choose lower-quality nursing homes (Li et al. 2015). Quality of nursing homes vary widely and can have meaningful impacts on important outcome such as short-term mortality (Cheng 2023), so these patterns of racial segregation and disparities paint a worrying picture when it comes to tackling racial gaps in healthcare. Finally, most residents are covered (at least in part) by insurance (most often Medicare or Medicaid), so differential distance to nursing homes is typically a much more important factor in residents' nursing home choice compared to differences in out-of-pocket prices.

2.1 Data

The primary data source for this paper is the Minimum Data Set 2.0 (MDS). All nursing homes that receive federal funding are required to fill out MDS assessment forms at regular intervals (42 CFR §483.20).[7,8,9] Data collected from the MDS assessments includes information on residents' demographics, cognitive status, communication and hearing patterns, vision patterns, mood and behavior patterns, psychosocial well-being, physical functioning and structural problems, continence issues, disease diagnoses (including ICD-9 codes), health conditions, oral health, nutrition, dental status, skin conditions, activity pursuit patterns, medications, special treatments and procedures, and discharge potential.

I supplement the MDS with data on nursing homes from other sources. This includes the Online

[7] The set of nursing homes receiving federal funding account for roughly 96 percent of all nursing homes (Grabowski, Gruber, and Angelelli 2008).

[8] Assessment forms must completed upon admission, at discharge (or death), quarterly in between, and whenever there is a significant change in status.

[9] MDS forms are typically filled out by a registered nurse (RN), or at least certified by one. Any willful misrepresentation in the MDS forms may result in penalties under the False Claims Act. This is not limited to upcoding and variables that affect reimbursements directly but also other variables related to resident well-being. This is because nursing homes "must provide services to attain or maintain the highest practicable physical, mental, and psychosocial well-being of each resident" (42 CFR §1395i–3) to be certified to receive federal funding. Hence, any misrepresentation pertaining to resident wellbeing may be interpreted as being related to misrepresentation connected to a requirement for federal funding, and thus falls under the False Claims Act. Moreover, several studies on the accuracy of MDS data have found it to be fairly reliable (Shin and Scherer, 2009).

Survey Certification and Reporting (OSCAR) surveys (which contain information such as nursing homes' ownership status and staffing levels), data on deficiency citations, and five-star ratings for nursing homes.[10] As quality measures, I focus mainly on RN and LPN staffing as well as deficiency citations. Nursing homes are cited for deficiencies by inspectors either during their annual visits, or during inspections that inspectors carry out in response to a complaint. I refer to these two types of deficiencies as standard and complaint deficiencies respectively; data on complaint deficiencies was only available from 2006 onwards. Probably the most well-known nursing home quality measure is the star rating provided by the CMS. However, this measure is only available from the end of 2008 onwards. For more details on the data used in this paper, see Appendix section A.

2.2 Sample and Summary Statistics

When possible, I use the entire sample of residents in the US. However, this is computationally infeasible for some of the analysis, particularly when studying residents' choice sets and nursing homes' admissions practices. In these cases, I focus on residents in Florida since the sample size for this state is relatively large for several reasons. In particular, there is substantial racial segregation and disparities across nursing homes in Florida, and there are roughly as many hispanic and black residents in Florida, which allows me to compare effects for different racial groups.

Table 1a shows summary statistics for nursing homes residents in the entire US, overall and separately by race. We observe that residents are typically white, female, advanced in age, and have less than a bachelor's degree. Moreover, most of them are admitted from acute care hospitals (and are thus likely to be short stay), 8 percent die within 90 days of admission, and 22 percent are already diagnosed with dementia at admission. Comparing characteristics of residents from different racial groups, the most notable differences are that white residents are on average older, more educated, and more likely to die within 90 days of admission, compared to black and hispanic residents. This last fact may seem somewhat surprising, but is largely attributable to minority residents being substantially younger (and hence, in better health) than white residents at admission, possibly because white residents have better resources which allow them to avoid going to nursing homes unless truly necessary. Table 1b shows that the same qualitative patterns hold for nursing home residents in Florida, except that while black residents outnumber hispanic residents more than two-to-one nationwide, there are roughly the same

[10] The OSCAR data is available from 2000 onwards from LTCFocus.org, which is maintained by Brown University Center of Gerontology and Healthcare Research. LTCFocus is sponsored by the National Institute on Aging (1P01AG027296) through a cooperative agreement with the Brown University School of Public Health. Data on deficiencies, Medicare cost reports, and five-star ratings are available from the CMS website.

number of black and hispanic nursing home residents in Florida.

Table 1: Summary Statistics for Residents

(a) All Residents

	All	White	Black	Hispanic
Age	77.22	78.29	70.84	72.43
	(12.88)	(12.05)	(15.41)	(15.47)
Female	0.624	0.634	0.580	0.554
	(0.484)	(0.482)	(0.494)	(0.497)
Married	0.335	0.346	0.231	0.338
	(0.472)	(0.476)	(0.422)	(0.473)
At Least Bachelor's Degree	0.106	0.114	0.0582	0.0401
	(0.308)	(0.318)	(0.234)	(0.196)
Admitted from Acute Care Hospital	0.861	0.859	0.865	0.868
	(0.346)	(0.348)	(0.341)	(0.338)
Death Within 90 Days	0.0814	0.0843	0.0705	0.0533
	(0.273)	(0.278)	(0.256)	(0.225)
Dementia	0.221	0.223	0.222	0.207
	(0.415)	(0.416)	(0.416)	(0.405)
Number of Residents	10,647,753	8,886,097	1,065,493	416,295

Notes: This table contains summary statistics for residents who had their first stays in a nursing home between 2000 and 2010.

(b) Residents in Florida

	All	White	Black	Hispanic
Age	77.77	78.74	69.75	76.65
	(12.22)	(11.43)	(15.40)	(12.92)
Female	0.604	0.609	0.563	0.598
	(0.489)	(0.488)	(0.496)	(0.490)
Married	0.371	0.386	0.259	0.332
	(0.483)	(0.487)	(0.438)	(0.471)
At Least Bachelor's Degree	0.119	0.131	0.0520	0.0669
	(0.324)	(0.337)	(0.222)	(0.250)
Admitted from Acute Care Hospital	0.903	0.901	0.904	0.921
	(0.296)	(0.299)	(0.295)	(0.270)
Death Within 90 Days	0.0627	0.0652	0.0599	0.0407
	(0.242)	(0.247)	(0.237)	(0.198)
Dementia	0.208	0.207	0.187	0.249
	(0.406)	(0.405)	(0.390)	(0.432)
Number of Residents	901,252	746,363	76,814	68,974

Notes: This table contains summary statistics for residents who had their first stays in a nursing home between 2000 and 2010.

Next, Table 2 shows summary statistics for nursing homes across the US in column 1, and in Florida in column 2 (weighted by the number of residents admitted between 2008–2010). Nursing homes have over a hundred beds on average, and occupancy rates are often above 80 percent. In addition, more than half of nursing homes are owned by chains, and more than 60 percent are for-profit. The data also contains numerous measures of nursing home quality, included the number of deficiencies that nursing homes are cited for during annual inspections,[11] and staffing hours for registered nurses (RNs), Licensed Practical Nurses (LPNs), and Certified Nursing Assistants (CNAs). Nursing homes in Florida are reasonably representative of nursing homes across the US, although a higher percentage are for profit, and various quality measures (specifically, cited deficiencies and RN staffing) tend to be lower.

Table 2: Summary Statistics for Nursing Homes

	All	Florida
Number of Beds	126.3	129.5
	(89.69)	(56.40)
Occupancy Percentage	84.30	88.72
	(14.48)	(11.00)
Chain	0.573	0.571
	(0.495)	(0.495)
For Profit	0.635	0.776
	(0.481)	(0.417)
Cited Deficiencies	6.401	7.520
	(6.198)	(6.059)
Registered Nurse Staffing (hours per patient-day)	0.686	0.450
	(0.996)	(0.656)
Licensed Practical Nurse Staffing (hours per patient-day)	0.941	0.992
	(0.746)	(0.508)
Certified Nursing Assistant Staffing (hours per patient-day)	2.337	2.573
	(1.262)	(0.964)
Number of Nursing Homes	17,248	1,101

Notes: This table contains summary statistics for nursing homes, weighted by the number of residents admitted between 2000 and 2010.

[11]Specifically, in Table 2 I show standard deficiency citations. Data on complaint deficiencies (which nursing homes are cited for during inspections conducted in response to complaints) are only available from 2006 onwards.

2.3 Broad Patterns of Racial Segregation and Disparities

To measure racial segregation across nursing homes, I use the index of dissimilarity. This index measures segregation across two racial groups and lies between 0 and 1, representing perfect integration and complete segregation respectively. For two groups A, and B, the dissimilarity index D of some geographical region is defined by:

$$D = \frac{1}{2} \sum_j \left| \frac{a_j}{\sum_{j'} a_{j'}} - \frac{b_j}{\sum_{j'} b_{j'}} \right|,$$

where a_j (respectively b_j) is the number of residents of group A (B) in nursing home j. An interpretation of the index D is that it is the proportion of one of the two groups that would have to move to different nursing homes in order for the distribution of the groups in each nursing home to match the overall distribution of these groups in the geographical region we are considering. Since the index of dissimilarity is only defined for two distinct racial groups, I compute this measure separately based on black versus non-black racial groups, and hispanic versus non-hispanic racial groups.

Focusing on the distribution of state-level nursing home segregation shown in Figure 1, we observe that dissimilarity indices for most states range from 0.3 to 0.7, consistent with previous research finding that nursing home segregation are similar to those for residential segregation (Mack et al. 2020). Unsurprisingly, given the definition of the dissimilarity index, black/non-hispanic segregation and hispanic/non-hispanic segregation are positively correlated, as shown in Appendix Figure A.1.

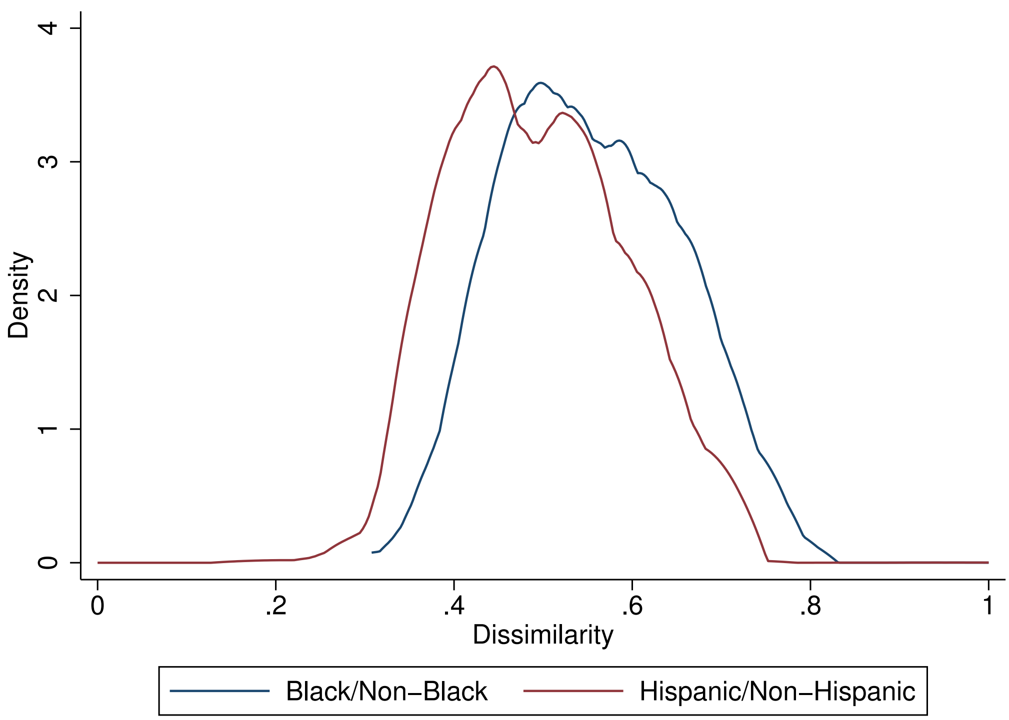
To measure racial disparities across nursing homes, I regress quality of the nursing home that residents are admitted to on race dummies, taking the coefficient estimates on the race dummies β^q_{black} and $\beta^q_{hispanic}$ as the racial gaps:

$$q_{j(i)} = \beta^q_0 + \beta^q_{black} black_i + \beta^q_{hispanic} hispanic_i + \epsilon^q_i. \tag{1}$$

Figure 2 shows the racial gap differs substantially across states, but a general pattern emerges whereby black residents tend to stay in nursing homes that have less registered nurse (RN) staffing, whereas the hispanic-white gap is much smaller.[12] Appendix Figures A.2, A.3, A.4, and A.5 show that the same patterns hold qualitatively when we use LPN staffing, fewer standard deficiencies, fewer complaint deficiencies, or star ratings as the quality measure.

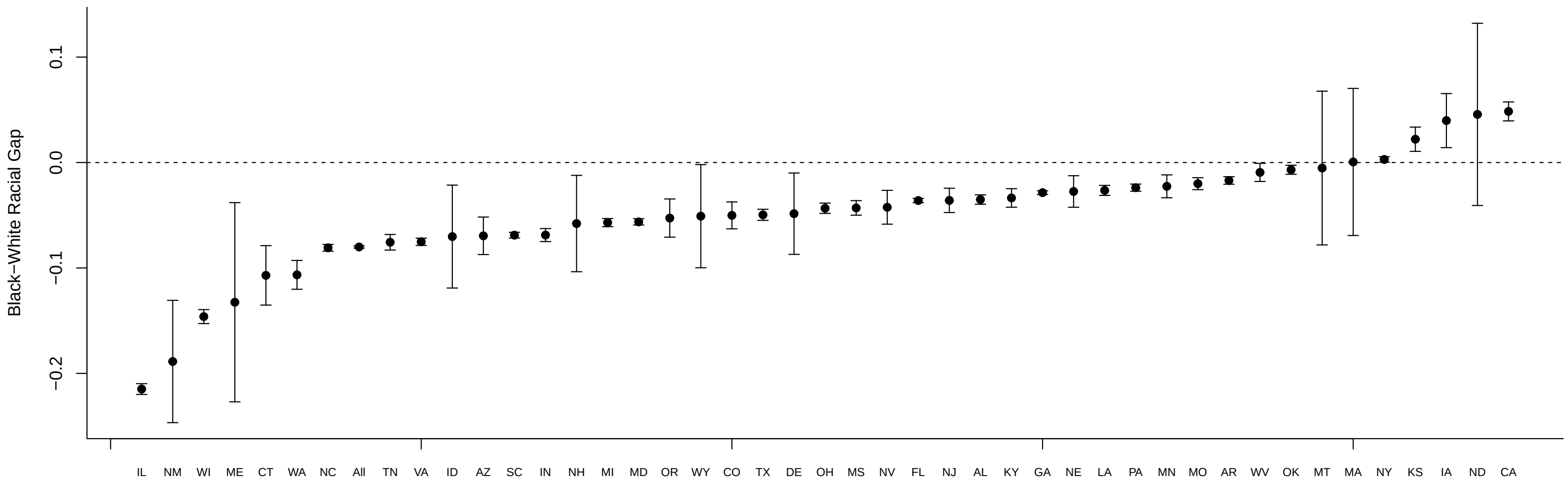
[12] Coefficient estimates with large standard errors due to small sample sizes are omitted for legibility purposes.

Figure 1: Patterns of Racial Segregation (Statewide Index of Dissimilarity)

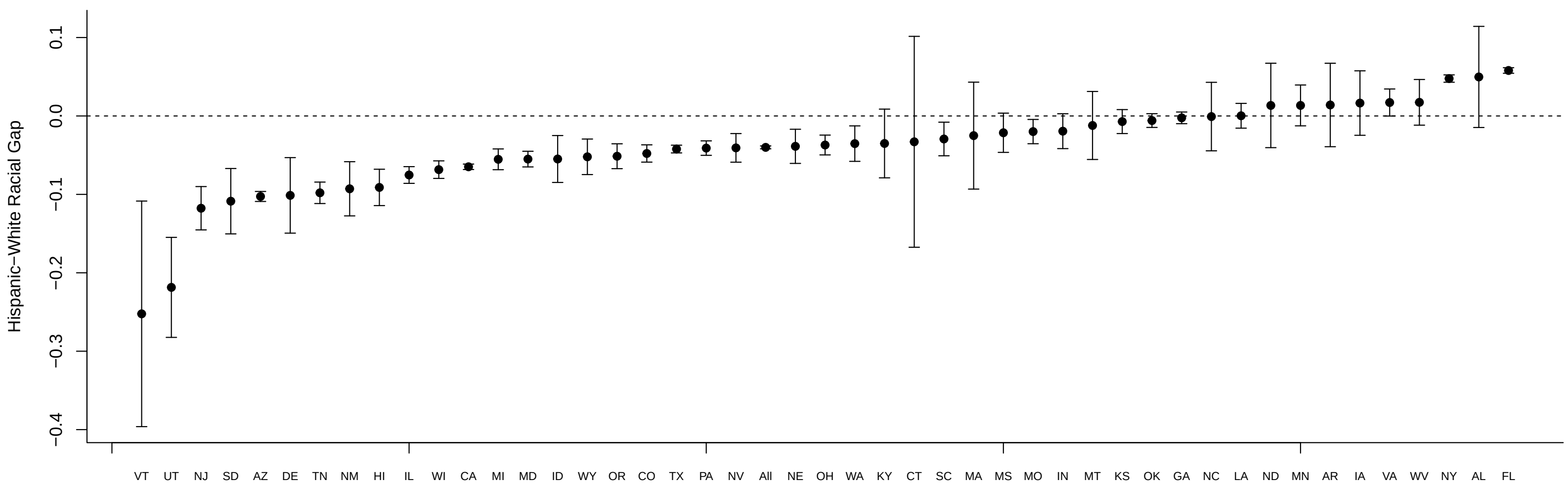


Notes: This figure shows kernel density plots of the dissimilarity index for black versus non-black residents and hispanic versus non-hispanic residents, measured at the state level.

Figure 2: Racial Gaps in Nursing Home Quality as Measured by RN Staffing Levels



(a) Black-White Gap



(b) Hispanic-White Gap

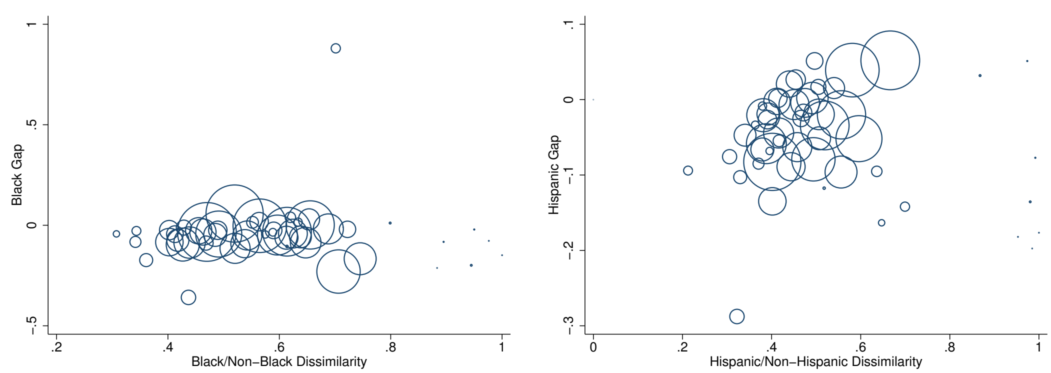
Notes: These figures display the estimated racial gaps in nursing home quality by state. Error bars indicate 95 percent confidence intervals for the estimates.

However, despite substantial statewide racial segregation and disparities, the cross-sectional relationship between the two are weak. Indeed, the scatterplots in Figure 3 shows that there is at best a weak cross-sectional relationship between the estimated racial disparities (based on RN staffing) and segregation at the state level. This absence of a clear cross-sectional relationship persists when we consider racial disparities based on other measures, such as LPN staffing, fewer standard deficiencies, and fewer complaint deficiencies, as Appendix Figures A.6, A.7, and A.8 show. This may seem surprising at first glance, but the relationship between racial segregation and disparities is complicated and far from mechanical: while disparities (as measured by differences in the quality of nursing homes that residents of different races are admitted to) must be zero when there is perfect integration, it is also possible for there to be zero disparities in a scenario of complete segregation. Better understanding the relationship will require us to dig deeper into the underlying mechanisms for racial segregation and disparities, which we do in the following sections.

Figure 3: Cross-Sectional Relationship Between State-Level Segregation and Disparities

(a) Black/Non-Black

(b) Hispanic/Non-Hispanic



Notes: These figures display scatter plots of the estimated racial gap (based on RN staffing) against segregation at the state level. Observations are weighted by the number of residents admitted to the state.

3 Reduced Form Evidence

In this section, I present several pieces of reduced form evidence on the potential causes of segregation and disparities across US nursing homes.

3.1 Nursing Home Segregation is Linked to Residential Segregation

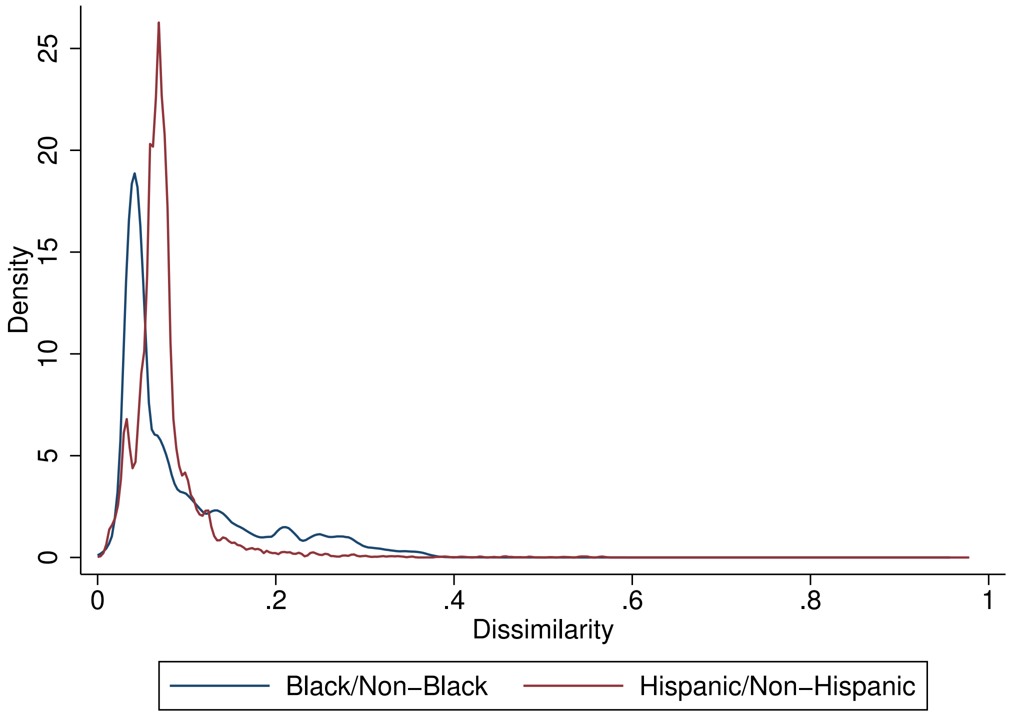
While the state-level dissimilarity indices plotted in Figure 1 show substantial levels of nursing home segregation similar to residential segregation, it is worth noting that the dissimilarity index depends on the geographical level at which it is measured. Specifically, if residents prefer nursing homes close to where they live, then in the presence of residential segregation, the distribution of racial groups within nursing homes will differ across different residential neighborhoods even in the absence of other disequalizing forces. The appropriate geographical level for measuring the dissimilarity index depends on the policy question one wishes to answer, but in order to isolate segregation arising from sources other than residential segregation one should compute the dissimilarity index at a geographical level that better approximates each resident's choice set.

Hence, given that most residents choose nursing homes within 15 miles of where they used to live, I compute dissimilarity indices based on a 15-mile radius of each 5-digit zip code for any resident's prior residential address. As a thought experiment, if distance to nursing homes does not matter for residents choosing their nursing homes and residents are willing to travel to any nursing home within their own state, then this local measure of segregation will be identical to the statewide measure. By contrast, if distance does matter to residents, and residential segregation is the only source of residential segregation, then the local segregation measure will be close to zero. In fact, Figure 1 shows the the truth lies somewhere in between these two extremes: dissimilarity indices are almost all smaller than 0.4, as compared to 0.3–0.7 at the state level, which suggests that residential segregation is an important explanation for overall nursing home segregation, but dissimilarity indices for many zip codes are also significantly different from zero, rejecting the notion that residential segregation is the sole cause.

3.2 Residential Segregation Alone is Unlikely to Explain Racial Disparities

Given the evidence that residential segregation feeds into nursing home segregation, there is a common perception this is also responsible for racial disparities, specifically that minorities live in poor neighborhoods where the quality of healthcare is low (Rahman and Foster 2015). However, I show in Table 3 that residential segregation is unlikely to fully explain disparities: conditional on zip code of prior residence, minority residents still tend to be admitted to nursing homes with lower staffing levels. Conditioning on zip code of residence typically reduces the estimated gap in quality of nursing homes minorities are admitted to, but a substantial gap still remains, and in fact it widens the black-white

Figure 4: Patterns of Racial Segregation (Index of Dissimilarity Based on 15-Mile Radius)



Notes: This figure shows kernel density plots of the dissimilarity index for black versus non-black residents and hispanic versus non-hispanic residents, measured based on neighborhoods in a 15-mile radius of each zip code and at the state level.

gap in the case of LPN staffing.[13] Appendix Table A.1 shows that the same qualitative patterns hold when using different deficiency measures and star ratings as the quality measure.

To better understand why residential segregation seems to have a more limited effect on disparities, in Appendix Table A.2 I examine the characteristics of nursing homes close to individuals of different races. We observe that differences in average registered nurse (RN) and licensed practitioner nurse (LPN) staffing as well as deficiency citations at nursing homes close to different racial groups are relatively small, and moreover, minority residents typically live closer to a larger number of nursing homes (a pattern consistent with the relatively higher proportion of minorities living in urban areas). Hence, access to higher-quality nursing homes in terms of geographical proximity alone cannot easily

[13]It is somewhat surprising that black residents go to nursing homes with higher LPN staffing compared to whites (unconditional on prior zip code of residence), but this is likely due to the fact that RN and LPN staffing are often seen as substitutes, so in many cases black residents may be going to nursing homes with higher LPN staffing levels but lower RN staffing levels than white residents.

Table 3: Association Between Nursing Home Staffing Levels and Minority Status

	RN Staffing		LPN Staffing	
	(1)	(2)	(3)	(4)
Race: Black	-0.0807***	-0.0418***	0.0219***	-0.0235***
	(0.000554)	(0.000682)	(0.000701)	(0.000880)
Race: Hispanic	-0.0425***	-0.0282***	-0.0195***	-0.0225***
	(0.000971)	(0.00115)	(0.00129)	(0.00155)
Constant	0.392***	0.388***	0.845***	0.850***
	(0.000220)	(0.000216)	(0.000271)	(0.000269)
Zip Code Fixed Effects		X		X
Number of Observations	8,577,363	8,575,899	8,568,306	8,566,842
R-squared	0.002	0.107	0.000	0.108

Notes: The unit of observation is a resident. Robust standard errors are shown in parentheses.

explain the substantial racial gap.

3.3 Discrimination by Nursing Homes May Give Rise to Disparities

Geographical proximity is not the only potential barrier to access: nursing homes may also discriminate against minorities in their admissions process. In fact, Gandhi (2020) shows that when capacity is strained, nursing homes tend to become more selective in the types of residents they admit and are less likely to admit Medicaid residents (who tend to be less profitable), presumably because the option value of an empty bed is increasing in capacity strain. Nursing homes may also find minority residents less attractive if certain minority characteristics are negatively correlated with profitability (e.g., Medicaid status) or due to outright taste-based discrimination. If this is the case and higher-quality nursing homes experience greater demand, then these selective admissions practices may give rise to the observed racial disparities.

To probe this possibility, I test two predictions from Gandhi's model. First, due to capacity constraints, nursing homes should admit fewer new residents when they are close to capacity. Second, and more importantly, the characteristics of residents that nursing homes admit during times of high and low occupancy should differ, given that nursing homes are more selective when they are closer to capacity.

To test the first prediction, I run regressions at the nursing home-day level of the form:

$$admissions_{jd} = \alpha^{cap} + \beta^{cap} occupancy_{jd} + \delta_{jm}^{cap} + \epsilon_{jd}^{cap},$$

where $admissions_{jd}$ and $occupancy_{jd}$ are measures of new admissions and occupancy of nursing home j on day d respectively, controlling for nursing-home month fixed effects δ_{jm}^{cap} in order to isolate temporary occupancy fluctuations (as opposed to longer term expansions and contractions in capacity). The results in Table 4a indicate that indeed, nursing homes admit fewer residents when occupancy is higher than usual, and Appendix Table A.3 shows that these results are robust to different measures of new admissions.

I test the second prediction by running regressions at the resident level of the form:

$$x_{ip} = \alpha_p^{select} + \beta_p^{select} occupancy_{j(i),d(i)} + \gamma_{\sim p}^{select} x_{i,\sim p} + \delta_{j,p}^{select} + \epsilon_{ip}^{select},$$

where x_{ip} is some characteristic p of resident i, $occupancy_{j(i),d(i)}$ is a measure of the nursing home when it admitted i, controlling for nursing home fixed effects $\delta_{j,p}^{select}$, and either controlling for other resident characteristics $x_{i,\sim p}$ or not in different specifications.

The results in Tables 4b and 4c indicate that when occupancy is higher than usual, nursing homes are less likely to admit minority and Medicaid residents, but are more likely to admit post-acute care residents (who tend to be covered by Medicare, which has higher reimbursement rates than Medicaid). The lower admission rates for minorities during times of capacity strain seen in Table 4b is consistent with discriminatory admissions practices against minorities, and the fact that the general pattern persists even after controlling for characteristics other than race suggests that some of this discrimination may be taste-based (as opposed to profit-maximizing statistical discrimination). Appendix Table A.4 shows that the same qualitative patterns generally hold using other measures of nursing home occupancy: nursing homes are less likely to admit black and Medicaid residents and more likely to admit hispanic residents when capacity is strained, although the results for hispanic residents are less clear.

3.4 In-Group Preferences May Explain the Perpetuation of Disparities

Another potential explanation for why minorities continue choosing lower-quality nursing homes despite the presence of higher-quality nursing homes nearby is in-group preferences (Schelling 1971). For example, individuals from a given racial group may prefer to interact with other members of the same racial group due to shared experiences, or because they believe they will be treated with more respect. If this is the case, then an initial distribution of minorities concentrated in lower-quality nursing homes

Table 4: Admissions Behavior by Nursing Homes

(a) Evidence of Capacity Constraints

	Number of New Residents		
	(1)	(2)	(3)
Lagged 7-Day Avg. Log Occupancy	-0.513***		
	(0.0458)		
Lagged 7-Day Avg. Occupancy		-0.00885***	
		(0.000347)	
Lagged 7-Day Avg. Occ. Percentile			-0.00309***
			(0.000122)
Nursing Home-Month Fixed Effects	X	X	X
N	2,345,772	2,345,772	2,345,772

Notes: This table shows regression results at the nursing home-day level wherein the dependent variable is number of new patients, and the independent variables are various measures of nursing home occupancy. Standard errors are clustered at the nursing home level.

(b) Evidence of Selective Admissions (Unconditional)

	Black	Hispanic	Medicaid	Post-Acute
	(1)	(2)	(3)	(4)
Lagged 7-Day Avg. Log Occupancy	-0.0283*	-0.0288**	-0.125***	0.0992***
	(0.0154)	(0.0123)	(0.0110)	(0.0160)
Nursing Home Fixed Effects	X	X	X	X
Number of Residents	666,278	666,278	666,278	666,278
R-squared	0.111	0.063	0.038	0.112

Notes: Regressions are at the resident level. Standard errors are clustered by nursing home.

(c) Evidence of Selective Admissions (Conditional)

	Black	Hispanic	Medicaid	Post-Acute
	(1)	(2)	(3)	(4)
Lagged 7-Day Avg. Log Occupancy	-0.0200	-0.0248**	-0.101***	0.0524***
	(0.0151)	(0.0121)	(0.0106)	(0.0156)
Nursing Home Fixed Effects	X	X	X	X
Controls for Other Characteristics	X	X	X	X
Number of Residents	666,278	666,278	666,278	666,278
R-squared	0.166	0.116	0.105	0.194

Notes: Regressions are at the resident level, and include controls for race, Medicaid, post-acute care, dementia, age, gender, marital status, and education (as long as the variable is not the dependent variable). Standard errors are clustered by nursing home.

may persist moving forward even in the absence of other inequities.

A key prediction of models of in-group preferences is that a shock to the minority share may lead to an equilibrium switch: specifically, a positive shock to the share of a racial group r at a nursing home may have persistent effects, as future residents of race r find that nursing home more attractive while residents of other races r' find it less attractive (Card, Mas and Rothstein 2007; Billings, Demming and Rockoff 2014; Hailey 2022). While the sample sizes for individual nursing homes are too small to conduct a tipping point analysis of Card, Mas and Rothstein in the neighborhood context based on Census tract data,[14] here I present evidence using dynamic panel methods and event study-type analyses.

I start by estimating autoregressive models at the nursing home-year level, based on the share of admitted residents who are of a minority group (either black or hispanic):

$$s^r_{jt} = \alpha^{r,ingroup} + \beta^{r,ingroup} s_{j,t-1} + \delta^{r,ingroup}_j + \gamma^{r,ingroup}_{ct} + \epsilon^{r,ingroup}_{jt}, \tag{2}$$

where s^r_{jt} is the share of residents admitted to nursing home j that are of race r, and observations are weighted by the number of admissions the nursing home j receives in year t. I control for time-invariant factors affecting share of black admissions by including nursing home fixed effects $\delta^{r,ingroup}_j$, and demographic trends by including county-year fixed effects $\gamma^{r,ingroup}_{ct}$.

The OLS estimates of equation (2) in column 1 of panels A and B in Table 5 show that a 10 percentage point (pp.) higher share of black (respectively, hispanic) admissions for a nursing home in a given year predicts 2.5pp. (1.8pp.) higher black (hispanic) admissions in the following year. The inclusion nursing home fixed effects and the relatively short panel (with only 10 years of data) may raise concerns about the Nickell bias (1981), so in columns 2 and 3 I estimate specifications based on dynamic panel methods from Anderson and Hsiao (1982) and Arellano and Bond (1991). The results for these estimation are similar to the OLS estimates qualitatively, although the exact magnitude differs across specifications. Finally, as a robustness check, Appendix Table A.5 shows that we obtain the same patterns when using numbers of minority residents instead of shares.

As a second test of the equilibrium switch behavior predicted by models of in-group preferences, I estimate event studies of the share of minority residents admitted in a given year, using large positive shocks to the share of minority residents as the event. Specifically, I estimate event studies of the form:

$$s^r_{jt} = \alpha^{r,shock} + \sum_{k \in \{-F,L\} \setminus \{-1\}} \beta^{r,shock}_k \mathbb{I}[t - E^r_j = k] + \delta^{r,shock}_j + \gamma^{r,shock}_t + \epsilon^{r,shock}_{jt},$$

where E^r_j is the year in which nursing home j receives a much higher than usual number of minority

[14] Specifically, the model predicts an unstable equilibrium for the minority share in a neighborhood, and a perturbation in the minority share above (respectively, below) this point may lead to a stable equilibrium with the neighborhood being filled mainly with minority (majority) individuals.

Table 5: Reduced Form Evidence of In-Group Preferences

Panel A: Share of Admitted Residents who are Black			
	OLS	*Anderson-Hsiao*	*Arellano-Bond*
	(1)	(2)	(3)
Black Share (Previous Admits)	0.246***	0.143***	0.0359***
	(0.0133)	(0.0229)	(0.0112)
Nursing Home Fixed Effects	X	Differenced-out	Differenced-out
County x Year Fixed Effects	X	X	X
Number of Nursing Home-Years	114,962	100,608	112,017
Panel B: Share of Admitted Residents who are Hispanic			
	OLS	*Anderson-Hsiao*	*Arellano-Bond*
	(1)	(2)	(3)
Hispanic Share (Previous Admits)	0.184***	0.120***	0.0412**
	(0.0183)	(0.0344)	(0.0172)
Nursing Home Fixed Effects	X	Differenced-out	Differenced-out
County x Year Fixed Effects	X	X	X
Number of Nursing Home-Years	114,962	100,608	112,017

Notes: This table shows regression results at the nursing home-year level, with weights equal to the number of residents who were admitted to the nursing home during that year. The Anderson-Hsiao and Arellano-Bond specifications correspond to dynamic panel methods from Anderson and Hsiao (1982) and Arellano and Bond (1991) respectively. Standard errors are clustered at the nursing home level.

residents of race r, i.e., $s^r_{j,E^r_j} - s^r_{j,E^r_j-1} \geq C$ for some threshold C. Mechanically, $\beta_0^{r,shock}$ will be large given the way the event is defined; instead, the test of in-group preferences is whether this shock leads to persistently higher shares of minority admission in future years, i.e., whether $\beta_k^{r,shock} > 0$ for $k > 0$.[15]

Figures 6a and 6b show results from these event studies using methods from Borusyak, Jaravel, and Spiess (2021), where I consider a year-to-year increase of at least 25pp. in the black and hispanic shares of admissions respectively. Consistent with in-group preferences, we observe that a positive shock to the black (hispanic) share of admissions of at least 25pp. leads to a more than 10pp. higher share of black (hispanic) share of admissions in each of the following 5 years. We observe that there is little evidence of pretrends preceding for black admissions other than the dip at $t = -1$ (which is mechanical, and a consequence of how the event is defined), and while some of the pretrend coefficients are statistically significant for hispanic admissions, the magnitude of these coefficients are much smaller

[15]Note that I use share of minority admissions rather than composition of current residents as the outcome, so that any persistent increase after the shock is not mechanical (as it may have been if I instead used composition of current residents as the outcome if some residents admitted in previous period(s) remain the nursing home for a long duration of time).

than the effect size. In Appendix Figure A.9, I conduct the same exercise but define the event as either a 10, 15, or 20pp. increase in the black or hispanic share of admissions as the "event". We observe qualitatively similar results: with a shock to minority share of admissions having persistent effects, and relatively little evidence of pretrends.

3.5 Information Frictions May Also Explain Disparities

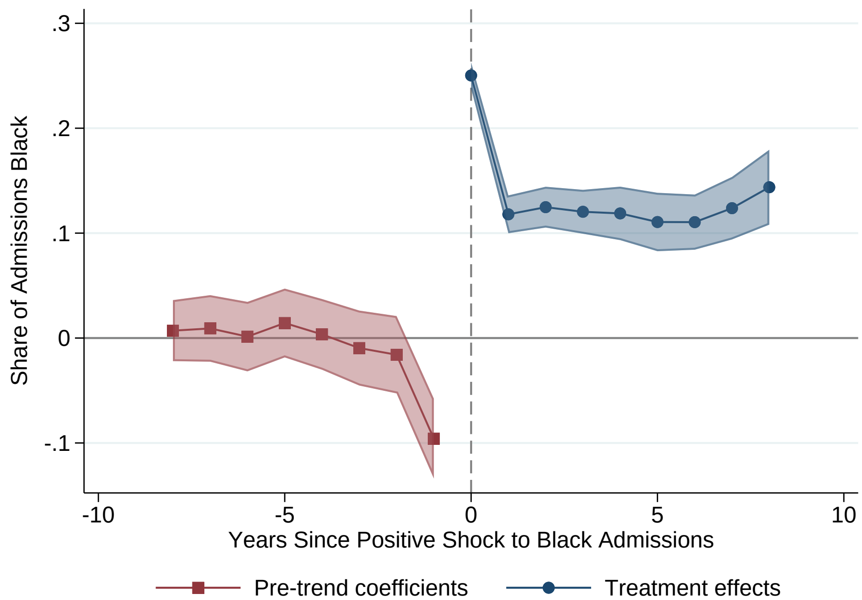
While in-group preferences may explain how racial disparities may persist over time, this explanation does not shed light on how these racial disparities arose in the first place. Other than discriminatory admissions practices, another possibility is that minorities may face greater barriers in accessing information about nursing home quality. Indeed, information about nursing home quality is readily available on the Nursing Home Compare website maintained by the CMS, and yet a large body of work has shown that internet use is substantially lower among racial minorities (e.g., Jain et al. 2021). Moreover, Cheng (2023) estimates that demand for quality among minority nursing home residents in California is lower than for white residents.

As a test of whether information frictions may explain racial disparities, I exploit the introduction of the five-star ratings system for nursing homes by the CMS at the end of 2008. If minorities do indeed face greater information frictions, we may expect smaller improvements in their nursing home choices relative to white residents. Specifically, I run difference-in-differences regressions of the form:

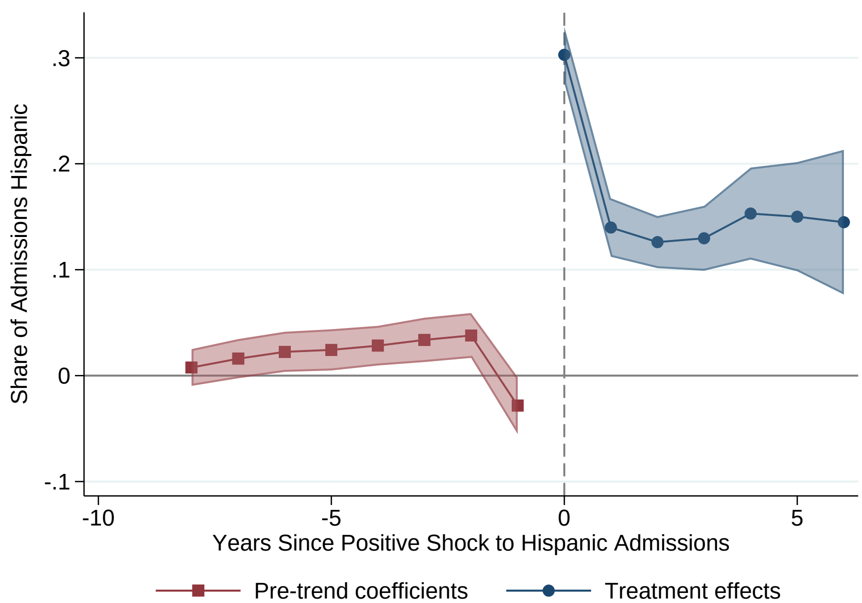
$$\bar{q}_{j(i)} = \alpha_0^{did} + \alpha_1^{did} black_i + \alpha_2^{did} hispanic_i + \alpha_3 post_i + \beta_{black}^{did} black_i \times post_i + \beta_{hispanic}^{did} hispanic_i \times post_i + \epsilon_i^{did},$$

where $post_i$ is an indicator for whether the resident was admitted to a nursing home after the introduction of star ratings, and $\bar{q}_{j(i)}$ is a time-invariant measure of the quality of the nursing home that resident i was admitted to. I use a time-invariant quality measure in order to abstract away from endogenous quality adjustments or gaming of the quality measure by nursing homes in response to the introduction of star ratings: for example, if lower-quality nursing homes inflate their staffing numbers in response to the star ratings, then it may seem like minorities are choosing better nursing homes even if their behavior does not change. For most quality measures, I use nursing home quality as of 2007 (prior to the star ratings), except for the star rating itself since it was only introduced at the end of 2008. If the parallel trends assumption holds, the coefficients β_{black}^{did} and $\beta_{hispanic}^{did}$ compares the extent to which minorities' nursing home choices improve following the introduction of star ratings, relative

Figure 5: Event Study on the Effect of a Positive Shock to the Share of Minority Admissions



(a) Increase in Share of Black Admissions



(b) Increase in Share of Hispanic Admissions

to white residents. To test the parallel trends assumption, I also estimate event study specifications of the form:

$$\bar{q}_{j(i)} = \alpha_r^{star} + \gamma_t^{star} + \sum_{k \neq 2008} \left(\beta_{black}^{star} \mathbb{I}[t(i) = k] \times black_i + \beta_{hispanic}^{star} \mathbb{I}[t(i) = k] \times hispanic_i \right) + \epsilon_i^{star},$$

where α_r^{star} and γ_t^{star} are race and time fixed effects respectively.

The difference-in-differences estimates in Table 6 suggests that nursing home choice for minorities seems to improve by a smaller extent than for white residents after the star ratings were introduced: while the estimates are not always statistically significant they are mostly negative, especially for hispanic residents. However, event study estimates in Appendix Figure A.10 are mostly statistically insignificant, and in some cases display pretrends. Hence, the reduced form evidence on information frictions can only be regarded as suggestive.

Table 6: Difference-in-Differences Estimates of the Effect of Star Ratings on Nursing Home Choice

	RN Staffing (2007)	*LPN Staffing (2007)*	*Fewer Standard Deficiencies (2007)*	*Fewer Complaint Deficiencies (2007)*	*Star Ratings (2009)*
	(1)	(2)	(3)	(4)	(5)
Race: Black	-0.0712***	0.0237**	-0.572***	-0.762***	-0.316***
	(0.00794)	(0.0109)	(0.140)	(0.136)	(0.0276)
Race: Hispanic	-0.0474***	-0.0462***	-0.539**	0.0415	0.0459
	(0.0120)	(0.0134)	(0.218)	(0.140)	(0.0532)
Post (2009-2010)	0.0342***	0.0654***	0.377***	0.280***	0.0664***
	(0.0108)	(0.0166)	(0.0346)	(0.0309)	(0.00633)
Post x Black	-0.0150	-0.0190	0.0727	0.0746	-0.0279**
	(0.0126)	(0.0176)	(0.0588)	(0.0552)	(0.0110)
Post x Hispanic	-0.0289***	-0.0345***	-0.113	-0.115*	-0.0445**
	(0.0101)	(0.0123)	(0.0957)	(0.0649)	(0.0202)
Number of Observations	4,902,133	4,893,699	4,342,081	2,619,654	4,997,402
R-squared	0.002	0.001	0.002	0.004	0.006

Notes: This table shows difference-in-differences estimates of the effect of the introduction of star ratings at the end of 2008 on the quality of nursing homes that residents of different races are admitted to, based on data from 2005-2010. Standard errors are clustered at the nursing home level.

A more fundamental difficulty with the reduced form evidence on potential causes of racial segregation and disparities presented in this section is that it is difficult to disentangle demand and supply side explanations. For example, consider the finding that improvements in the quality of nursing homes that minorities are admitted to were smaller compared to white residents following the introduction of star ratings. This could either be due to greater information frictions faced by minorities, or because higher-quality nursing homes are now in greater demand and thus reject minority applicants at higher rates. Moreover, it is hard to quantify the extent to which each of these explanations contribute to segregation and disparities based on reduced form evidence alone. Hence, in the next section, I introduce a structural model to address these issues.

4 Structural Estimation

4.1 Overview of Empirical Matching Model

In order to disentangle residents' preferences from nursing homes' admission decisions, I estimate an empirical matching model similar to Agarwal and Somaini (2022) and Cheng (2023). I assume that resident i's indirect decision utility for each nursing home $j \in \mathcal{J}_i \equiv \{j | dist_{ij} \leq 15 \; miles\}$ is given by:

$$v_{ij} = w_j' \kappa^1 + w_j' \kappa^2 x_i + dist_{ij}' \kappa^{dist} + \epsilon_{ij}, \tag{3}$$

where x_i and w_j are resident and nursing home characteristics respectively, $dist_{ij}$ is a measure of distance between resident i and nursing home j, and ϵ_{ij} is an idiosyncratic utility shock. For the location normalization, I omit the constant term, and to set the scale normalization, I assume that $\epsilon_{ij} \sim N(0, 1)$.

Nursing homes' admissions policies are given by:

$$\pi_{ij} = x_i' \psi^1 + w_j' \psi^2 x_i + occ_{d(i)j}' \psi^{occ} + \omega_{ij}, \tag{4}$$

where $occ_{d(i)j}$ is a measure of nursing home j's occupancy in the period leading up to i's admission date $d(i)$, and ω_{ij} is an idiosyncratic shock. I assume that nursing home j is willing to admit resident i if and only if $\pi_{ij} \geq \underline{\pi}$, so resident i's constrained choice set is $\{j \in \mathcal{J}_i | \pi_{ij} \geq \underline{\pi}\}$. Note that these constraints are not observed in the data, hence the reason we need to estimate this empirical matching model. The location normalization is set by including a constant term in equation (4) and setting $\underline{\pi} = 0$, and the scale normalization is achieved by assuming $\omega_{ij} \sim N(0, 1)$.

To elaborate on how this model incorporates elements such as in-group preferences, information frictions, and discrimination, equation (3) typically takes the form:

$$\begin{aligned} v_{ij} =& \kappa_0^{black} s_{d(i)j}^{black} + \kappa_0^{hisp} s_{d(i)j}^{hisp} + \kappa_1^{black} s_{d(i)j}^{black} black_i + \kappa_1^{hisp} s_{d(i)j}^{hisp} hispanic_i \\ &+ q_j' \kappa_0^q + black_i q_j' \kappa_{black}^q + hispanic_i q_j' \kappa_{hisp}^q + dist_{ij}' \kappa^{dist} + \epsilon_{ij}, \end{aligned}$$

where $s_{d(i)j}^r$ is the share of nursing home j's admissions that are of race r in the 365 days leading up to i's admission date $d(i)$. We can think of κ_0^r and $\kappa_0^r + \kappa_1^r$ respectively as measuring preferences among those of race $r' \neq r$ and race r for a higher share of recently admitted residents being of race r. In the absence of racial preferences among residents, we will expect $\kappa_0^r = \kappa_1^r = 0$. Similarly, κ_0^q captures white residents' demand for quality, whereas $\kappa_0^q + \kappa_r^q$ captures demand for quality among residents of race r (for non-white residents). If we assume that there is no "true" racial heterogeneity in demand for quality and no information frictions, then we should have $\kappa_r^q = 0$. In the simplest specification for the supply side, I estimate:

$$\pi_{ij} = \psi_0 + \psi^{black} black_i + \psi^{hisp} hispanic_i + \tilde{x}_i' \psi^{\tilde{x}} + occ_{d(i)j}' \psi^{occ} + \omega_{ij},$$

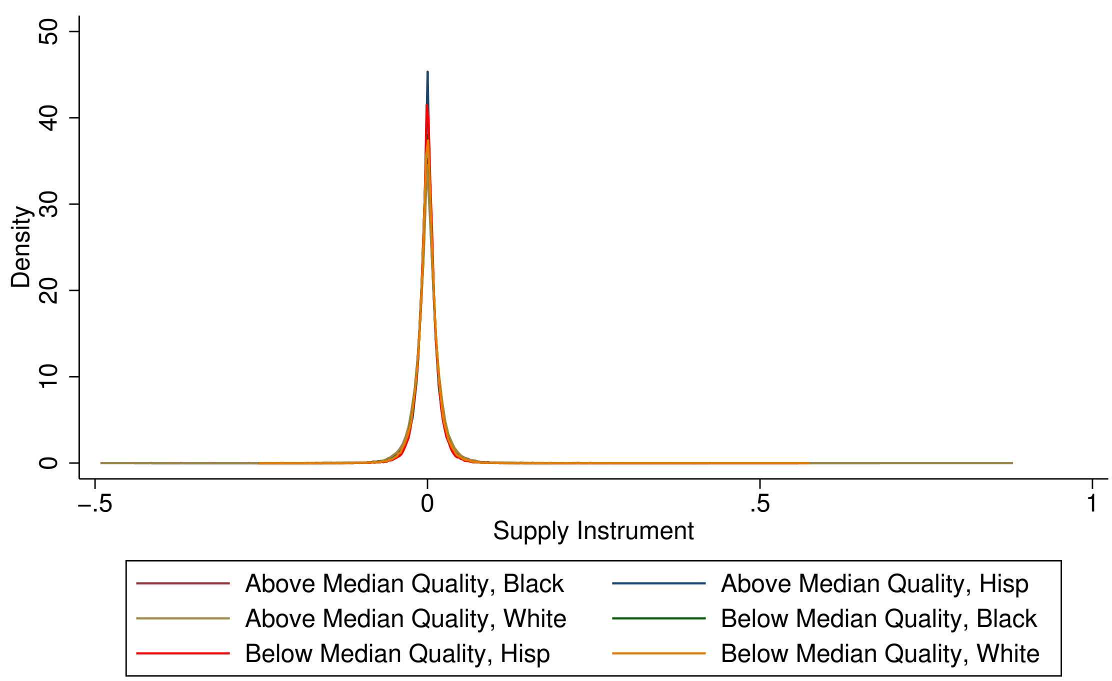
where $\tilde{x}_i$ are some none-race characteristics of resident i. In the absence of discriminatory admissions practices, we would expect $\psi^r = 0$.

Agarwal and Somaini (2022) provide a sharp set of identification conditions for such a model, and the key substantive requirement is that we need both demand and supply side instruments. I use distance as the demand side instrument and temporary fluctuations in nursing homes' occupancy (specifically log occupancy residualized of nursing home-month fixed effects) as the supply instrument. The relevance condition for both instruments are likely to be satisfied: residents have a strong preference for nursing homes close to where they used to live, and nursing homes are less likely to admit new residents when capacity is strained, as seen in Table 4a. The exclusion restriction for the demand instrument is also likely satisfied, since nursing homes are unlikely to be care about where their residents originally lived.

The exclusion restriction for the supply instrument deserves closer scrutiny, and for better intuition we start by considering why using occupancy (instead of temporary fluctuations in occupancy) is likely to violate the exclusion restriction, and its implications for structural model's estimates. All else, equal residents may prefer less crowded nursing homes (thus violating the exclusion restriction), and if higher-quality nursing homes are in greater demand, this will lead us to underestimate residents' demand for quality. Our estimates of racial heterogeneity in demand for quality (specifically, the gap in demand between white and minority residents) may still remain valid, but will no longer remain so if preferences for "crowdedness" also varies by race.

The use of *temporary fluctuations* in nursing home occupancy (specifically, within nursing home-month fluctuations) addresses this concern in two ways. First, short-term fluctuations in occupancy are less likely to matter for residents. Second, by residualizing the occupancy measure of nursing home-month fixed effects, we eliminate potential correlations between occupancy and nursing home characteristics such as quality. This is illustrated in Figure 6, which shows that the distribution of temporary occupancy fluctuations of nursing homes within 15 miles of each resident is essentially identical across above-median and below-median-quality nursing homes (as measured by RN staffing) close to white, black, and hispanic residents. Appendix Figure A.11 shows that the same pattern holds if we consider other quality measures.

Figure 6: Exclusion Restriction for Temporary Occupancy Fluctuations



Notes: This figure display kernel density plots of temporary occupancy fluctuations (defined as lagged 7-day log occupancy residualized of nursing home-month fixed effects) of nursing homes within 15 miles of each resident at their time of admission.

Finally, estimation of this model has to deal with the curse of dimensionality. In particular, there are $2^{|\mathcal{J}_i|} - 1$ possible constrained choice sets for resident i, so methods such as maximum likelihood that require us to sum over each distinct possibility are computationally infeasible. Hence, I use Gibbs sampling with data augmentation (for v_{ij} and π_{ij}) for my estimation, since this obviates the need to individually compute the probability of each potential choice set, without needing to make additional substantive assumptions.

At a high level, in each iteration of the Gibbs sampler, I draw utility and profit shocks ϵ_{ij} and ω_{ij} in such a way that the resulting latent variables respect the matching outcomes. I then update the posterior distribution of the parameters before moving onto the next iteration. Under regularity conditions, the draws of the parameters will eventually converge to their stationary distribution, and we can form Bayesian confidence sets based on the distribution of these draws, which are also endowed with a frequentist interpretation as a consequence of the Berstein von-Mises theorem. Due to computational constraints, I limit the sample to residents and nursing homes Florida between 2008–2010. For details on the algorithm for the Gibbs sampler, see Appendix section B.

4.2 Structural Estimation Results

Table 7 shows results from my estimation of the empirical matching model. The estimates of the demand and supply instruments are all highly statistically significant, which is reassuring for the identification of the model.

Column 1 shows that non-black residents tend to prefer nursing homes that admitted fewer black residents recently whereas black residents prefer the opposite, and the same holds for hispanic and non-hispanic residents, consistent with in-group preferences. In addition, the estimates indicate that minority residents are less responsive to nursing home quality (as measured by fewer complaint deficiencies) than white residents, although this pattern is more pronounced for black residents (and less so for hispanic residents). On the supply side, we observe that nursing homes are less likely to admit black and hispanic residents when they are close to capacity, which might suggests nursing home discrimination against black residents.

To account for preferences over unobservable time-invariant factors nursing home characteristics, I include nursing home fixed effects in residents' utility equation in column 2. Hence, demand for racial characteristics of recently admitted residents is identified based on variation over time, similar to the fixed effects regressions in Table 5 and event study in Figure 5. I do not include the level terms (for previous share of different races and quality) when I include nursing home fixed effects since there is too little year-to-year variation to estimate the coefficients for these terms with any meaningful precision. Nonetheless, we observe that the inclusion of nursing home fixed effects in residents' utility equation remain quite similar qualitatively to the specification without fixed effects.

The driving force behind estimates of preference heterogeneity shown above need not be race-based, in the sense that race is correlated with other characteristics (such as education) and that these other variables are more fundamental reasons for the preference heterogeneity. Similarly, the previous estimates do not give us any indication whether the nature of nursing homes' discrimination is taste-based or statistical. To shed some light on these issues, in columns 3 and 4 I control for residents' age, education, and Alzheimer's status. Specifically, I allow for in-group preferences based on these characteristics, for demand for quality to vary according to these variables, and for nursing homes' admissions policies to depend on these characteristics.

The estimates of in-group preferences and racial heterogeneity in sensitivity to quality remain largely unchanged qualitatively when we control for heterogeneity by other resident characteristics. On the other hand, we find smaller estimates of nursing homes' unwillingness to admit minority residents although these are sometimes still statistically significant at the 5 percent level. This is consistent with the presence of both taste-based and statistical discrimination by nursing homes against minority groups.

As a robustness check, in Appendix Table A.6, I include preferences for additional measures of nursing home quality, namely registered nurse (RN) staffing and licensed practitioner nurse (LPN) staffing, which I again allow to vary by race. The results consistently show that minority residents are less sensitive to different measures of nursing home quality, consistent with information frictions. In addition, the results on in-group preferences and discriminatory admissions practices remain qualitatively unchanged.

5 Counterfactuals

In this section, I conduct counterfactual simulations to assess how important different factors are for explaining segregation and choice disparities. Specifically, I use the structural estimates from the previous section to simulate the dynamic evolution of segregation and disparities by modifying different parameters. In these simulations, I abstract away from endogenous quality adjustments by nursing homes, setting quality for each nursing home to its average over the time period of the structural estimation (2008–2010).

Table 7: Estimates of Residents' Preferences and Selective Admissions by Nursing Homes

	(1)	(2)	(3)	(4)
Resident Preferences				
Utility (Distance to Facility)	-0.171***	-0.191***	-0.177***	-0.202***
	(0.009)	(0.008)	(0.008)	(0.007)
Utility (Previous Share Black)	-1.519***		-0.429***	
	(0.143)		(0.042)	
Utility (Black x Previous Share Black)	1.624***	0.517***	0.295***	0.539***
	(0.064)	(0.033)	(0.022)	(0.032)
Utility (Previous Share Hispanic)	-0.789***		-0.058*	
	(0.137)		(0.041)	
Utility (Hispanic x Previous Share Hispanic)	1.202***	0.561***	0.391***	0.608***
	(0.069)	(0.036)	(0.026)	(0.038)
Utility (Less Deficiencies)	-0.007***		-0.074***	
	(0.002)		(0.005)	
Utility (Less Deficiencies x Black)	-0.023***	-0.026***	-0.019***	-0.029***
	(0.003)	(0.003)	(0.002)	(0.003)
Utility (Less Deficiencies x Hispanic)	-0.018***	-0.001	-0.011***	-0.003*
	(0.003)	(0.003)	(0.003)	(0.003)
Nursing Homes' Admission Policies				
Occupancy	-5.843***	-5.609***	-4.412***	-4.282***
	(1.033)	(0.976)	(0.547)	(0.488)
Race (Black)	-0.833***	-1.013***	-0.174***	-0.304***
	(0.199)	(0.233)	(0.033)	(0.038)
Race (Hispanic)	-0.876***	-0.903***	-0.392***	-0.437***
	(0.177)	(0.194)	(0.064)	(0.052)
Profit Intercept	1.453***	1.472***	-1.578***	-2.422***
	(0.313)	(0.352)	(0.325)	(0.304)
Controls for Other Characteristics			X	X
Nursing Home Fixed Effects in Utility		X		X

Notes: This table shows estimates of the structural model using Gibbs sampling. A burn-in period corresponding to the first half of the chain was used. The controls for other characteristics in columns 3 and 4 refers to in-group preferences by education, age, and dementia status, heterogeneity in demand for quality by these characteristics, and including these characteristics in nursing homes' admissions policies.

To simulate a successful ban on discriminatory admissions practices nursing homes, I set the race-specific parameters ψ^r in the admissions equation to zero. To mimic the elimination of in-group preferences, I set the parameters for race $(\kappa_0^{r\prime}, \kappa_1^{r\prime})'$ in residents' utility equation to zero. Next, to simulate the elimination of information frictions (or an effective information intervention targeted at minorities), I set the interactions between nursing home quality and minority race dummies $\kappa_r^q = 0$.[16] Finally, to mimic the elimination of residential segregation, I randomize the zip code of prior residence for each resident, effectively randomizing them to different counterfactual choice sets. For more details on the simulations, see Appendix section C.

[16]This is a conservative definition of information frictions, since it implicitly assumes that white residents have full information. Cheng (2023) finds that average demand for quality among nursing home residents is very low (e.g., much lower than demand estimates from hospital settings) among most residents, suggesting that white residents likely also face some information frictions.

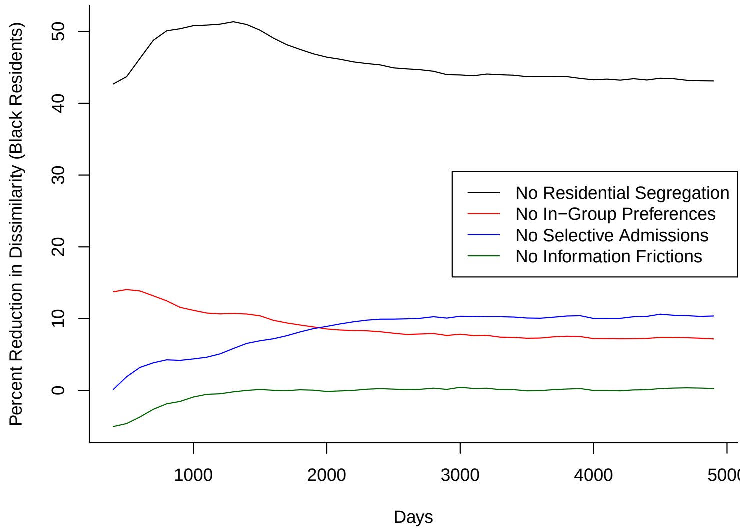
Figure 7 shows the effect of various counterfactual policies on segregation and choice disparities for black residents in Florida. I use the statewide dissimilarity index and coefficients from estimates of equation (1) to measure the evolution over time of racial segregation and disparities in the simulations. Specifically, I plot the average difference over 100 simulations between simulated segregation and disparities based on counterfactual parameter values (or randomized residential location) and those based on the original estimated parameters, scaled by the mean value of segregation or disparities in the simulations based on estimated parameter values. The simulated of disparities (before taking the difference between simulations based on counterfactual and estimated/original values and scaling) is shown in Appendix Figure A.12, with the simulations based on estimated/original values in black, and simulations based on counterfactual values in red.

We observe that residential segregation is by far the most important contributor to nursing home segregation, accounting for more than 40 percent of segregation, whereas each of the other explanations only explain 10 percent or less of segregation. By contrast, information frictions is the main driver of racial disparities in terms of complaint deficiencies, even though it is orthogonal to segregation. Discrimination, in-group preferences, and residential segregation each explains less than half of the gap in racial disparities compared to information frictions.

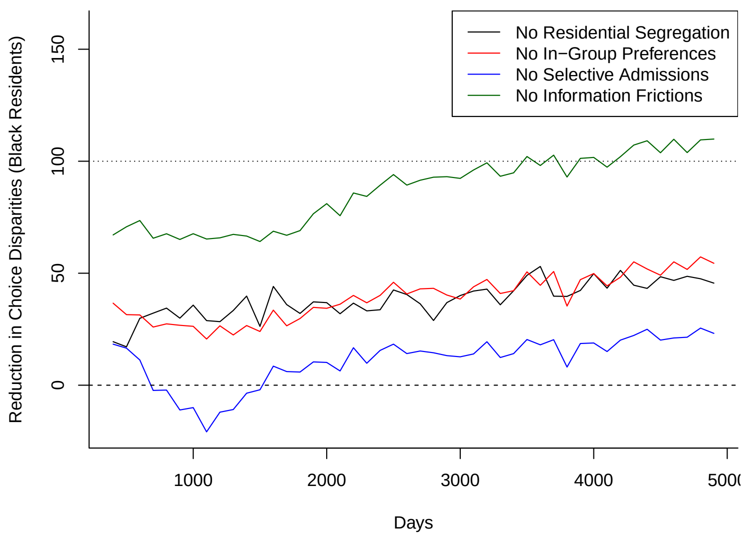
Finally, I also conduct these simulations using the structural estimates which incorporates several different measures of quality, and obtain qualitatively similar results, although the correlation between quality measures may make the results more difficult to interpret.[17] In particular, Appendix Figure A.13 shows that residential segregation is by far the most important explanation for black/non-black segregation, while information frictions is always the most important explanation for black-white disparities (or tied first with discrimination in the case of RN staffing). Interestingly, the results show that eliminating residential segregation may even widen racial disparities in terms of LPN staffing, although this can be rationalized by the observation in Appendix Table A.2 that average LPN staffing levels of nursing homes in black residents' choice set is in fact slightly higher than for white residents. Counterfactual simulation results for Hispanic residents are shown in Appendix Figure A.14, and are qualitatively similar on the segregation dimension, although the results on racial disparities are difficult to interpret given that average quality of nursing homes that white and hispanic residents are admitted to in Florida are quite similar in the data.

[17] In particular, RN and LPN staffing are often seen as substitutes, so reducing disparities in RN staffing may sometimes have an opposite effect of LPN staffing.

Figure 7: Simulated Effect of Counterfactual Policies on Racial Segregation and Disparities (Black Residents): Structural Model with Only Fewer Complaint Deficiencies as the Quality Measure



(a) Reduction in Segregation



(b) Reduction in Disparities (Fewer Complaint Deficiencies)

Notes: These figures the simulated effect of various counterfactuals on racial segregation and disparities for residents in Floridian nursing homes over time.

6 Conclusion

In this paper, I studied the question of why racial minorities are disproportionately concentrated in low-quality nursing homes. I find that the common notion that residential segregation is the main cause holds a grain of truth, but is far from the entire story. In particular, while it is certainly responsible for a large share of racial segregation across nursing homes, the same cannot be said for racial disparities. Instead, information frictions seem to be the main contributor to racial disparities, even though it explains little of racial segregation. In addition, discriminatory admissions practices and in-group preferences also play some role in explaining disparities.

References

[1] Agarwal and Somaini, 2022. "Demand Analysis under Latent Choice Constraints." Working paper.

[2] Anderson, Theodore Wilbur, and Cheng Hsiao, 1982. "Formulation and estimation of dynamic models using panel data." *Journal of Econometrics*, 18(2): 47–82.

[3] 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.

[4] Baicker, Katherine, and Amitabh Chandra, 2004. "Medicare Spending, The Physician Workforce, And Beneficiaries' Quality Of Care: Areas with a high concentration of specialists also show higher spending and less use of high-quality, effective care." *Health Affairs*, 23(Suppl1): W4–184.

[5] Billings, Stephen B., David J. Deming, and Jonah Rockoff, 2014. "School segregation, educational attainment, and crime: Evidence from the end of busing in Charlotte-Mecklenburg." *The Quarterly Journal of Economics*, 129(1): 435–476.

[6] Black, Sandra E, 1999. "Do better schools matter? Parental valuation of elementary education." *The Quarterly Journal of Economics*, 114(2): 577–599.

[7] Borusyak, Kirill, Xavier Jaravel, and Jann Spiess, 2022. "Revisiting event study designs: Robust and efficient estimation." Working paper.

[8] Card, David, Alexandre Mas, and Jesse Rothstein, 2008. "Tipping and the Dynamics of Segregation." *The Quarterly Journal of Economics*, 123(1): 177–218.

[9] Cheng, Alden, 2023. "Demand for Quality in the Presence of Information Frictions: Evidence from the Nursing Home Market." Working paper.

[10] Derenoncourt, Ellora, 2022. "Can you move to opportunity? Evidence from the Great Migration." *American Economic Review*, 112(2): 369–408.

[11] Gandhi, Ashvin, 2020. "Picking your patients: Selective admissions in the nursing home industry." Working paper.

[12] Hackmann, Martin B., 2019. "Incentivizing better quality of care: The role of Medicaid and competition in the nursing home industry." *American Economic Review*, 109(5): 1684–1716.

[13] Hailey, Chantal A., 2022. "Racial preferences for schools: Evidence from an experiment with White, Black, Latinx, and Asian parents and students." *Sociology of Education*, 95(2): 110–132.

[14] Hurd, Michael D., Pierre-Carl Michaud, and Susann Rohwedder, 2017. "Distribution of lifetime nursing home use and of out-of-pocket spending." *PNAS*, 104(37): 9838–9842.

[15] Li, Yue, Charlene Harrington, Helena Temkin-Greener, Kai You, Xueya Cai, Xi Cen, and Dana B. Mukamel, 2015. "Deficiencies in care at nursing homes and racial/ethnic disparities across homes fell, 2006–11." *Health Affairs*, 34(7): 1139–1146.

[16] Mack, Deborah S., Bill M. Jesdale, Christine M. Ulbricht, Sarah N. Forrester, Pryce S. Michener, and Kate L. Lapane, 2020. "Racial segregation across US nursing homes: A systematic review of measurement and outcomes." *The Gerontologist*, 60(3): e218–e231.

[17] Margo, Robert A., 1991. "Segregated schools and the mobility hypothesis: A model of local government discrimination." *The Quarterly Journal of Economics*, 106(1): 61-73.

[18] Rahman, Momotazur, and Andrew D. Foster, 2015. "Racial segregation and quality of care disparity in US nursing homes." *Journal of Health Economics*, 39: 1–16.

[19] Schelling, Thomas C., 1971. "Dynamic models of segregation." *Journal of Mathematical Sociology*, 1(2): 143–186.

Appendix

A Data Appendix

Data on residents was obtained from the Minimum Data Set 2.0 (MDS), which was then superceded by the Minimum Data Set 3.0 after 2010. I focus on the earlier period given inconsistencies in the variables collected across versions, and the fact that the earlier version contains residents' zip code of prior residence. In principle, zip code for some residents in the later period can be obtained by linking the MDS data with Medicare and Medicaid data, but I will still have to drop residents without Medicare or Medicaid. Data on nursing homes was obtained from the Online Survey Certification and Reporting (OSCAR) data, and was downloaded from LTCFocus, a product of the Shaping Long-Term Care in America Project being conducted at the Brown University Center for Gerontology and Healthcare Research and supported, in part, by the National Institute on Aging. This data contains yearly level information on nursing homes such as the street address, and average RN, LPN, and CNA staffing levels. Additional information on deficiency citations and star ratings were obtained from the CMS website.

In constructing my sample of residents, I drop the relatively small number of residents with errors in birth or death dates (e.g., with different birth or death dates recorded across different assessments). Race for black and hispanic individuals is coded based on the corresponding variable in the MDS data, and the base category (which I refer to as white in the main text) includes a small number of Asian and individuals of other races. For the structural demand estimation sample, I consider only resident-nursing home pairs within 15 miles of each other. In addition, I drop nursing homes that admit fewer than 30 residents over the period of my structural estimation sample (2008–2010) to ensure sufficient power.

Two key variables for the structural estimation are distance between residents and nursing homes (which serves as the demand-side instrument), and temporary occupancy fluctuations (which I use as the supply-side instrument). To construct distances, I combine residents' zip code from the MDS, with nursing homes' street address from the OSCAR data, which I convert to latitude and longitude using the Google Maps API. The stata command "geodist" was then used to compute the distance between residents' zip code of prior residence, and nursing homes' locations.

For temporary occupancy fluctuations, I consider the average log occupancy (or average occupancy or within-nursing home occupancy percentile as robustness checks) of nursing homes within 15 miles of each resident in the week before the resident was admitted to the nursing home. I focus on the week prior to admissions instead of a shorter term measure such as the day of/before admission given that residents (or hospital discharge planners) typically need some time to search for and coordinate with nursing homes. In addition, I residualize lagged 7-day log occupancy of nursing home-month fixed effects, to abstract from expansions or contractions that a nursing home may be undergoing (for example, if a nursing home is expanding, it will be more likely to admit new residents even though its occupancy seems high). The OSCAR contains data on total number of beds at a nursing home which one could in principle use as a measure of capacity, but this variable is measured with substantial error and only updated annually. In fact, measures of occupancy over time based on residents' admission and discharge dates from the MDS frequently exceeds total number of beds reported by nursing homes in the OSCAR data, which calls into question the use of this variable as a measure of capacity.

B Algorithm for Gibbs Sampler

In the following description for the Gibbs sampler, when drawing structural error terms in sequence for $j \in \mathcal{J}_i$, I assume an increasing order (although obviously any other order works as well). In addition, to simplify notation, I denote variables in residents' utility and nursing homes' admissions equations by $\mathbf{X}_{ij}$ and W_{ij} respectively,[18] and refer to the nursing home that resident i ends up in by $\mu(i)$.

Denoting iterations of the Gibbs sampler by k and indicating the values of various parameters in the kth iteration of the Gibbs sampler using a superscript k, the steps for implementing the Gibbs sampler are as follows.

1. Initialization ($k = 0$): I assume that $(\epsilon_{ij}, \omega_{ij}) \sim^{i.i.d.} N(0, I_2)$ and set the following conjugate priors for the parameters: $(\kappa', \psi')' \sim N(0, 100I)$.

    (a) Set the initial values of the parameters $\theta^0 = \left(\kappa^{0\prime}, \psi^{0\prime}\right)$ at their prior mean.

    (b) Initial data augmentation: For each resident i, draw the vector ϵ_i^0 such that $v^0_{i,\mu(i)} \geq v^0_{ij}$ for all $j \in \mathcal{J}_i$.

        i. Draw $\omega^0_{i,\mu(i)}$ such that $\omega^0_{i,\mu(i)} \geq -W'_{ij}\psi^0$ and for $j \neq \mu(i)$ draw ω^0_{ij} from the unconditional distribution.

[18] These include resident characteristics x_i, nursing home characteristics w_j, distance between residents and nursing homes $dist_{ij}$, occupancy fluctuations at nursing homes occ_{ij}, and interactions between these variables.

ii. Set $\epsilon^0_{i,\mu(i)}$ equal to three times the standard deviation of the prior. For $j \neq \mu(i)$, draw ϵ^0_{ij} such that $\epsilon^0_{ij} \leq \left(\mathbf{X}_{i,\mu(i)} - \mathbf{X}_{ij}\right)' \kappa^0 + \epsilon^0_{i,\mu(i)}$ if $\pi^0_{ij} \geq 0$ or draw ϵ^0_{ij} unconditionally otherwise.

2. For $k+1 = 1, ..., K$:

(a) Draw the profit shocks $\omega^{k+1}_i | \upsilon^k_i; \psi^k$ in sequence for $j \in \mathcal{J}_i$.

i. If $\upsilon^k_{ij} < \upsilon^k_{i,\mu(i)}$, draw ω^{k+1}_{ij} unconditional on assignment (given that even if i is eligible for j, i would not choose j).

ii. If $\upsilon^k_{ij} > \upsilon^k_{i,\mu(i)}$, draw ω^{k+1}_{ij} from a truncated normal with mean and variance given by the conditional distribution and truncation point $\omega^{k+1}_{ij} < -W_{ij}\psi^k$ (given that otherwise i would choose j over $\mu(i)$).

iii. Finally, if $j = \mu(i)$, draw from the conditional distribution with truncation point given by $\omega^{k+1}_{ij} \geq -W'_{ij}\psi^k$ (given that i must always be eligible for the facility she was ultimately assigned to).

(b) Update π^{k+1}_i according to $\pi^{k+1}_{ij} = W'_{ij}\psi^k + \omega^{k+1}_{ij}$.

(c) Draw the utility shocks $\epsilon^{k+1}_i | \pi^{k+1}_i; \kappa^k$ in sequence, for $j \in \mathcal{J}_i$.

i. If $\pi^{k+1}_{ij} < 0$, draw ϵ^{k+1}_{ij} unconditionally (given that i would not choose such a facility even if she were eligible for it).

ii. If $\pi^{k+1}_{ij} \geq 0$ and $j \neq \mu(i)$, draw ϵ^{k+1}_{ij} from the conditional distribution with truncation point given by $\upsilon^{k+1}_{ij} < \mathbf{X}'_{ij}\kappa^k$.

iii. For $j = \mu(i)$, draw $\epsilon^{k+1}_{i,\mu(i)}$ such that $\upsilon^{k+1}_{i,\mu(i)}$ is larger than the current values of $\upsilon_{i,j'}$ for $j' \neq j$ and $\pi_{ij'} \geq 0$.

(d) Update υ^{k+1}_i according to $\upsilon^{k+1}_{ij} = \mathbf{X}'_{ij}\kappa^k + \epsilon^{k+1}_{ij}$.

(e) Update the parameters θ based on the new indirect utilities υ^{k+1} and profits π^{k+1}.

i. First, we update κ. Denote the design matrix in the equation for indirect utilities by $\mathbb{X}$. In matrix notation, we have:

$$\upsilon = \mathbb{X}\kappa + \epsilon, \ \epsilon \sim N(0, I).$$

We have a normal conjugate prior for κ, with mean μ^0_κ and covariance matrix Σ^0_κ. The posterior distribution of κ conditional on v and W is:

$$\kappa|(v, \mathbb{X}) \sim N(\tilde{\mu}_\kappa, \tilde{\Sigma}_\kappa),$$

where the posterior mean and covariance matrix are given by:

$$\begin{aligned}
\tilde{\mu}_\kappa &= \left(\frac{\mathbb{X}'\mathbb{X}}{\sigma_\epsilon^2} + \left(\Sigma_\kappa^0\right)^{-1}\right)^{-1} \left(\left(\Sigma_\kappa^0\right)^{-1} \mu_\kappa^0 + \frac{\mathbb{X}'\kappa}{\sigma_\epsilon^2}\right) \\
&= \left(\mathbb{X}'\mathbb{X} + \left(\Sigma_\kappa^0\right)^{-1}\right)^{-1} \left(\left(\Sigma_\kappa^0\right)^{-1} \mu_\kappa^0 + \mathbb{X}'\kappa\right),
\end{aligned}$$

$$\begin{aligned}
\tilde{\Sigma}_{\theta_v} &= \left(\frac{\mathbb{X}'\mathbb{X}}{\sigma_\epsilon^2} + \left(\Sigma_\kappa^0\right)^{-1}\right)^{-1} \\
&= \left(\mathbb{X}'\mathbb{X} + \left(\Sigma_\kappa^0\right)^{-1}\right)^{-1}.
\end{aligned}$$

We then set κ^{k+1} by drawing from this posterior distribution.

A. Next, we will update ψ. Denote the design matrix in the equation for the admissions rule by W. In matrix notation, we have:

$$\pi = W\psi + \omega, \ \omega \sim N(0, I).$$

We have a normal prior for ψ, with mean μ_ψ^0 and covariance matrix Σ_ψ^0, so the posterior distribution of θ_π conditional on π and W is:

$$\psi|(\pi, W) \sim N(\tilde{\mu}_\psi, \tilde{\Sigma}_\psi),$$

with posterior mean and covariance matrices given by:

$$\begin{aligned}
\tilde{\mu}_\psi &= \left(\frac{W'W}{\sigma_\omega^2} + \left(\Sigma_\psi^0\right)^{-1}\right)^{-1} \left(\left(\Sigma_\psi^0\right)^{-1} \mu_\psi^0 + \frac{W'\psi}{\sigma_\omega^2}\right) \\
&= \left(W'W + \left(\Sigma_\psi^0\right)^{-1}\right)^{-1} \left(\left(\Sigma_\psi^0\right)^{-1} \mu_\psi^0 + W'\psi\right),
\end{aligned}$$

$$
\begin{aligned}
\tilde{\Sigma}_{\psi} &= \left(\frac{W'W}{\sigma_{\omega}^{2}} + \left(\Sigma_{\psi}^{0} \right)^{-1} \right)^{-1} \\
&= \left(W'W + \left(\Sigma_{\psi}^{0} \right)^{-1} \right)^{-1}.
\end{aligned}
$$

We then set ψ^{k+1} by drawing from this posterior distribution.

C Simulation Details

Recall that our structural model is based on equations for residents' decision utility and nursing homes' admission rules respectively:

$$
\begin{aligned}
v_{ij} = &\kappa_0^{black} s_{ij}^{black} + \kappa_0^{hisp} s_{ij}^{hisp} + \kappa_1^{black} s_{ij}^{black} black_i + \kappa_1^{hisp} s_{ij}^{hisp} hispanic_i \\
&+ q_j' \kappa_0^q + black_i q_j' \kappa_{black}^q + hispanic_i q_j' \kappa_{hisp}^q + dist_{ij}' \kappa^{dist} + \epsilon_{ij},
\end{aligned}
$$

$$
\pi_{ij} = \psi_0 + \psi^{black} black_i + \psi^{hisp} hispanic_i + \tilde{x}_i' \psi^{\tilde{x}} + occ_{ij}' \psi^{occ} + \omega_{ij}.
$$

Nursing home j is willing to admit resident i if and only if $\pi_{ij} \geq 0$, and resident i chooses the nursing home which yields the highest decision utility among the set of nursing homes that are willing to admit her. For computational feasibility, for each resident i, I only consider nursing homes within 15 miles of her $\mathcal{J}_i \equiv \{j | dist_{ij} \leq 15 \ miles\}$. I denote estimated using "hats", e.g., $(\hat{\kappa}', \hat{\psi}')'$, but for the counterfactuals I switch to using "stars".

To simulate the elimination of in-group preferences, I set $\kappa_0^{r*} = \kappa_1^{r*} = 0$, and to simulate the elimination of information frictions, I set $\kappa_r^{q*} = 0$. Similarly, to simulate the elimination of discriminatory admissions practices, I set $\psi^{r*} = 0$. To simulate the elimination of residential segregation, I permute the zip codes of prior address for residents. Hence, counterfactual distances between resident i and different nursing homes $dist_{ij}^*$ will generally differ from the original distances $dist_{ij}$, and i is faced with a different potential choice set $\mathcal{J}_i^* \equiv \{j | dist_{ij}^* \leq 15 \ miles\}$, unless she is randomized to the same zip code. By virtue of the permutation process, the unconditional geographical distribution of residents' prior addresses remains unchanged, but the distribution of races within each zip code will reflect the overall distribution of race (in expectation), hence eliminating residential segregation.

In terms of notation, I will use κ^*, ψ^*, and $dist^*$ throughout the description of simulations, and it should be understood that this is either equal to its estimated or original value if the corresponding component of the simulation is not turned on, or equal to a counterfactual value otherwise. For example, if we are considering a counterfactual with no residential segregation, $dist^*_{ij}$ will generally be different from $dist_{ij}$, whereas in a counterfactual where we take residential segregation as given, I still use the same notation $dist^*_{ij}$, but this will be equal to $dist_{ij}$.

The simulation algorithm is as follows:

1. Setup:

   (a) For each nursing home j, I set the share of j's admissions that are of race r in the last 365 days prior to the start of the simulation to the mean of this value over the estimation period.

   (b) For each nursing home j, I assume that its quality measures are time-invariant over the simulation period, and set this equal to the average over the estimation period.

   (c) For the simulations, I set the number of new arrivals each day to the mean in the data, which is $N^*_d = 177$.

2. Simulation: for day $d^* = 1, ..., D^* = 5000$ of the simulation:

   (a) If the counterfactual assumes no residential segregation, I permute the zip codes of prior address for residents.

   (b) I then randomly select $N^*_{d^*}$ residents and simulate their choices. For resident $i^* = 1, ..., N^*_d$:

      i. I draw $\epsilon_{ij} \sim N(0, 1)$ and compute:

$$\begin{aligned} v_{i^*j} =& \kappa_0^{black*} s_{d^*j}^{black*} + \kappa_0^{hisp*} s_{d^*j}^{hisp} + \kappa_1^{black*} s_{d^*j}^{black} black_i + \kappa_1^{hisp} s_{d^*j}^{hisp} hispanic_i \\ &+ q_j' \kappa_0^{q*} + black_i q_j' \kappa_{black}^{q*} + hispanic_i q_j' \kappa_{hisp}^{q*} + dist_{ij}'^{*} \kappa^{dist*} + \epsilon_{ij}, \end{aligned}$$

      for each nursing home $j \in \mathcal{J}^*_{i^*} \equiv \{j' | dist^*_{ij'} \leq 15 \; miles\}$.

      ii. Also, for each $j \in \mathcal{J}^*_{i^*}$, I draw $\omega_{i^*j} \sim N(0, 1)$ and compute:

$$\pi_{i^*j} = \psi_0^* + \psi^{black*} black_{i^*} + \psi^{hisp*} hispanic_{i^*} + \tilde{x}_{i^*}' \psi^{\tilde{x}*} + occ_{d(i^*)j}'^{*} \psi^{occ*} + \omega_{i^*j}.$$

      iii. I set i^*'s nursing home to be $\mu(i^*) \equiv argmax_j \{v_{i^*j} | \pi_{i^*j} \geq 0, j \in \mathcal{J}^*_{i^*}\}$.[19]

   (c) Next, I update the shares of residents admitted to each nursing home that is of each race in the 365 days leading up to the next day:

[19] If $\pi_{i^*j} < 0$ for all $j \in \mathcal{J}^*{}_{i^*}$, then I simply drop the resident, but this occurs extremely rarely in any of the simulations.

i. Letting N^r_{dj} be the number of residents of race r that is admitted to nursing home j on day d of the simulation and N_{dj} be the number of residents of any race that is admitted to j on day d, I set:

$$s^{r*}_{d^*+1,j} \equiv \frac{\sum_{d=d^*-364}^{d^*} N^r_{dj}}{\sum_{d=d^*-364}^{d^*} N_{dj}}.$$

(d) Finally, I update the occupancy measures of each nursing home for the next day in the simulation:

$$occ^*_{d^*+1,j} = log\left(\sum_{d=d^*-6}^{d^*} N_{dj}\right) - log(\bar{N}_j),$$

where $\bar{N}_j$ is the mean occupancy measure at nursing home j in the data.

3. Measuring segregation and disparities: to measure segregation and disparities on a day d^* for the simulations, I use data from the past 100 days (i.e., days $d^* - 99$ up to day d^*).

A Appendix Figures and Tables

Table A.1: Association Between Other Measures of Nursing Home Quality and Minority Status

	Fewer Standard Deficiencies		Fewer Complaint Deficiencies		2009 Star Ratings	
	(1)	(2)	(3)	(4)	(5)	(6)
Race: Black	-0.825***	-0.727***	-0.927***	-0.333***	-0.303***	-0.199***
	(0.00751)	(0.00865)	(0.00697)	(0.00437)	(0.00144)	(0.00160)
Race: Hispanic	-1.064***	-0.418***	0.0282***	-0.245***	0.0578***	-0.115***
	(0.0122)	(0.0133)	(0.00881)	(0.00580)	(0.00231)	(0.00238)
Constant	-6.846***	-6.882***	-1.995***	-0.985***	2.592***	2.589***
	(0.00233)	(0.00223)	(0.00188)	(0.00100)	(0.000479)	(0.000428)
Zip Code Fixed Effects		X		X		X
Number of Observations	8,578,937	8,577,473	4,218,959	8,577,473	8,578,937	8,457,229
R-squared	0.002	0.174	0.006	0.108	0.002	0.276

Notes: The unit of observation is a resident. Robust standard errors are shown in parentheses.

Table A.2: Characteristics of Nursing Homes in Residents' Choice Sets (By Race)

	All	White	Black	Hispanic
Nursing Homes in Choice Set	23.930	22.830	26.393	31.939
	(15.344)	(15.546)	(14.060)	(11.384)
Distance to Nursing Homes	7.897	7.858	7.616	8.551
	(1.978)	(2.019)	(1.984)	(1.333)
RN Staffing (for nursing homes in choice set)				
Mean	0.291	0.288	0.286	0.316
	(0.065)	(0.067)	(0.057)	(0.048)
Standard Deviation	0.158	0.156	0.155	0.185
	(0.066)	(0.068)	(0.059)	(0.046)
Minimum	0.082	0.085	0.076	0.056
	(0.057)	(0.058)	(0.053)	(0.045)
Maximum	0.705	0.694	0.700	0.817
	(0.366)	(0.378)	(0.329)	(0.234)
LPN Staffing (for nursing homes in choice set)				
Mean	0.946	0.942	0.955	0.970
	(0.115)	(0.118)	(0.108)	(0.084)
Standard Deviation	0.238	0.236	0.240	0.259
	(0.189)	(0.196)	(0.182)	(0.103)
Minimum	0.589	0.595	0.585	0.539
	(0.149)	(0.151)	(0.145)	(0.127)
Maximum	1.631	1.617	1.674	1.719
	(0.907)	(0.924)	(0.937)	(0.672)
Deficiencies (for nursing homes in choice set)				
Mean	0.656	0.659	0.718	0.551
	(0.844)	(0.846)	(0.928)	(0.714)
Standard Deviation	1.000	1.001	1.091	0.881
	(1.236)	(1.245)	(1.286)	(1.069)
Minimum	0.003	0.004	0.003	0.001
	(0.080)	(0.084)	(0.071)	(0.039)
Maximum	3.628	3.608	4.036	3.336
	(4.772)	(4.790)	(4.968)	(4.279)
Number of Residents	630,947	520,735	53,179	53,650

Notes: This table contains summary statistics for the choice sets of residents (defined as nursing homes within 15 miles of each resident) who had their first stays in a nursing home in Florida between 2000 and 2010.

Table A.3: Robustness Checks for Evidence on Capacity Constraints

(a) Dummy for Any New Admission as the Dependent Variable

	Any New Residents		
	(1)	(2)	(3)
Lagged 7-Day Avg. Log Occupancy	-0.314***		
	(0.0299)		
Lagged 7-Day Avg. Occupancy		-0.00504***	
		(0.000215)	
Lagged 7-Day Avg. Occ. Percentile			-0.00191***
			(7.10e-05)
Nursing Home-Month Fixed Effects	X	X	X
N	2,345,772	2,345,772	2,345,772

Notes: This table shows regression results at the nursing home-day level wherein the dependent variable is a dummy for any new residents, and the independent variables are various measures of nursing home occupancy. Standard errors are clustered at the nursing home level.

(b) Flow of Residents as the Dependent Variable

	Flow of Residents		
	(1)	(2)	(3)
Lagged 7-Day Avg. Log Occupancy	-4.767***		
	(0.342)		
Lagged 7-Day Avg. Occupancy		-0.0790***	
		(0.00148)	
Lagged 7-Day Avg. Occ. Percentile			-0.0278***
			(0.000636)
Nursing Home-Month Fixed Effects	X	X	X
N	2,345,772	2,345,772	2,345,772

Notes: This table shows regression results at the nursing home-day level wherein the dependent variable is the flow of residents (difference between number of residents today and yesterday), and the independent variables are various measures of nursing home occupancy. Standard errors are clustered at the nursing home level.

Table A.4: Robustness Checks for Evidence on Selective Admissions

(a) Using Occupancy in Levels as the Independent Variable (Unconditional)

	Black	Hispanic	Medicaid	Post-Acute
	(1)	(2)	(3)	(4)
Lagged 7-Day Avg. Occupancy	-0.000805***	0.000602***	-0.00146***	0.0127**
	(8.76e-05)	(0.000128)	(0.000193)	(0.00535)
Nursing Home Fixed Effects	X	X	X	X
Number of Residents	666,278	666,278	666,278	666,278
R-squared	0.111	0.063	0.038	0.112

Notes: Regressions are at the resident level. Standard errors are clustered by nursing home.

(b) Using Occupancy in Levels as the Independent Variable (Conditional)

	Black	Hispanic	Medicaid	Post-Acute
	(1)	(2)	(3)	(4)
Lagged 7-Day Avg. Occupancy	-0.000644***	0.000280**	-0.00136***	0.0153***
	(8.49e-05)	(0.000124)	(0.000186)	(0.00509)
Nursing Home Fixed Effects	X	X	X	X
Controls for Other Characteristics	X	X	X	X
Number of Residents	666,278	666,278	666,278	666,278
R-squared	0.111	0.063	0.038	0.112

Notes: Regressions are at the resident level, and include controls for race, Medicaid, post-acute care, dementia, age, gender, marital status, and education (as long as the variable is not the dependent variable). Standard errors are clustered by nursing home.

(c) Using Occupancy Percentile as the Independent Variable (Unconditional)

	Black	Hispanic	Medicaid	Post-Acute
	(1)	(2)	(3)	(4)
Lagged 7-Day Avg. Percentile	-0.000432***	0.000323***	-0.000730***	0.00666**
	(4.40e-05)	(6.43e-05)	(9.70e-05)	(0.00269)
Nursing Home Fixed Effects	X	X	X	X
Number of Residents	666,278	666,278	666,278	666,278
R-squared	0.111	0.063	0.038	0.112

Notes: Regressions are at the resident level. Standard errors are clustered by nursing home.

(d) Using Occupancy Percentile as the Independent Variable (Conditional)

	Black	Hispanic	Medicaid	Post-Acute
	(1)	(2)	(3)	(4)
Lagged 7-Day Avg. Percentile	-0.000348***	0.000152**	-0.000681***	0.00806***
	(4.26e-05)	(6.25e-05)	(9.36e-05)	(0.00256)
Nursing Home Fixed Effects	X	X	X	X
Controls for Other Characteristics	X	X	X	X
Number of Residents	666,278	666,278	666,278	666,278
R-squared	0.166	0.116	0.105	0.194

Notes: Regressions are at the resident level, and include controls for race, Medicaid, post-acute care, dementia, age, gender, marital status, and education (as long as the variable is not the dependent variable). Standard errors are clustered by nursing home.

Table A.5: Robustness Checks for Reduced Form Evidence of In-Group Preferences

Panel A: Number of Admitted Residents who are Black			
	OLS	*Anderson-Hsiao*	*Arellano-Bond*
	(1)	(2)	(3)
Number of Previous Black Admits	0.520***	2.563	0.145**
	(0.0347)	(1.803)	(0.0606)
Nursing Home Fixed Effects	X	Differenced-out	Differenced-out
County x Year Fixed Effects	X	X	X
Number of Nursing Home-Years	114,962	100,608	112,017
Panel B: Number of Admitted Residents who are Hispanic			
	OLS	*Anderson-Hsiao*	*Arellano-Bond*
	(1)	(2)	(3)
Number of Previous Hispanic Admits	0.515***	1.313***	0.532***
	(0.0193)	(0.304)	(0.0607)
Nursing Home Fixed Effects	X	Differenced-out	Differenced-out
County x Year Fixed Effects	X	X	X
Number of Nursing Home-Years	114,962	100,608	112,017

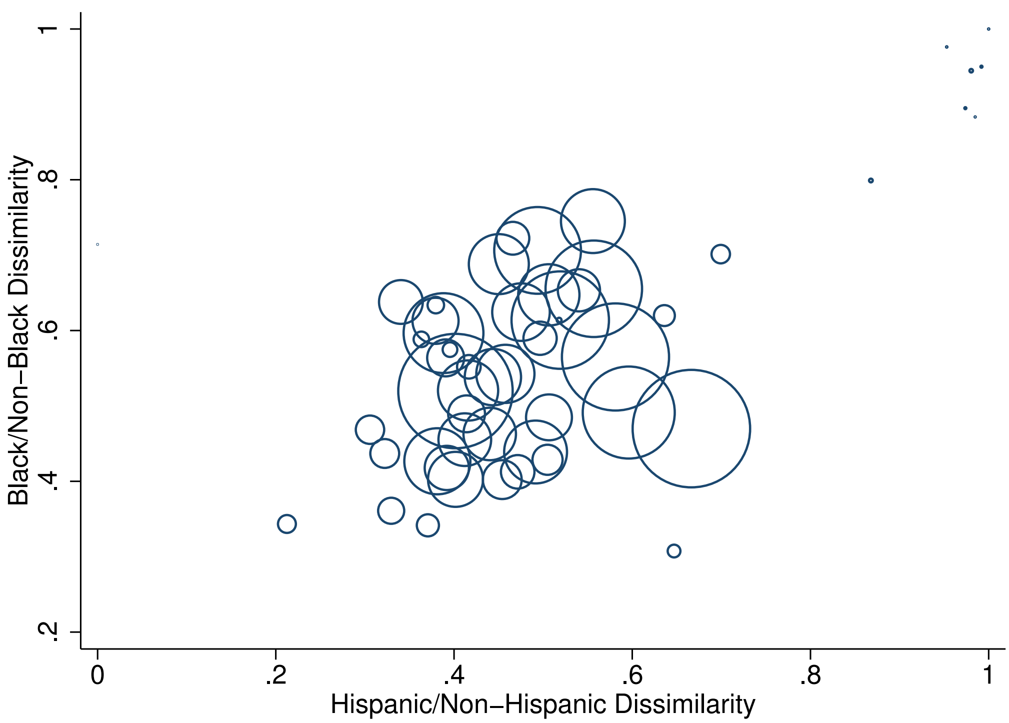
Notes: This table shows regression results at the nursing home-year level, with weights equal to the number of residents who were admitted to the nursing home during that year. The Anderson-Hsiao and Arellano-Bond specifications correspond to dynamic panel methods from Anderson and Hsiao (1982) and Arellano and Bond (1991) respectively. Standard errors are clustered at the nursing home level.

Table A.6: Estimates of Residents' Preferences and Selective Admissions by Nursing Homes

	(1)	(2)	(3)	(4)
Resident Preferences				
Utility (Distance to Facility)	-0.174***	-0.191***	-0.18***	-0.202***
	(0.006)	(0.007)	(0.006)	(0.007)
Utility (Previous Share Black)	-0.401***		-0.418***	
	(0.022)		(0.032)	
Utility (Black x Previous Share Black)	0.282***	0.503***	0.282***	0.523***
	(0.025)	(0.028)	(0.023)	(0.03)
Utility (Previous Share Hispanic)	-0.065***		-0.063**	
	(0.019)		(0.028)	
Utility (Hispanic x Previous Share Hispanic)	0.349***	0.54***	0.368***	0.585***
	(0.026)	(0.032)	(0.026)	(0.037)
Utility (Less Deficiencies)	-0.005**		-0.076***	
	(0.003)		(0.006)	
Utility (Less Deficiencies x Black)	-0.024***	-0.024***	-0.018***	-0.027***
	(0.002)	(0.002)	(0.002)	(0.003)
Utility (Less Deficiencies x Hispanic)	0.037	0.101***	0.092*	0.129***
	(0.051)	(0.036)	(0.053)	(0.041)
Utility (RN Staffing)	0.347***		0.406***	
	(0.033)		(0.069)	
Utility (RN Staffing x Black)	-0.203***	-0.126***	-0.2***	-0.122***
	(0.039)	(0.03)	(0.041)	(0.033)
Utility (RN Staffing x Hispanic)	0.037	0.101***	0.092*	0.129***
	(0.051)	(0.036)	(0.053)	(0.041)
Utility (LPN Staffing)	0.096*		0.406***	
	(0.068)		(0.069)	
Utility (LPN Staffing x Black)	-0.115***	-0.242***	-0.101***	-0.25***
	(0.03)	(0.03)	(0.031)	(0.03)
Utility (LPN Staffing x Hispanic)	-0.152***	-0.255***	-0.163***	-0.26***
	(0.044)	(0.031)	(0.044)	(0.035)
Nursing Homes' Admission Policies				
Occupancy	-5.756***	-5.453***	-4.351***	-4.255***
	(1.115)	(0.877)	(0.503)	(0.487)
Race (Black)	-0.678***	-1.001***	-0.169***	-0.276***
	(0.17)	(0.229)	(0.045)	(0.039)
Race (Hispanic)	-0.585***	-0.926***	-0.378***	-0.433***
	(0.124)	(0.211)	(0.071)	(0.05)
Profit Intercept	1.329***	1.471***	-1.565***	-2.422***
	(0.317)	(0.344)	(0.235)	(0.305)
Controls for Other Characteristics			X	X
Nursing Home Fixed Effects in Utility		X		X

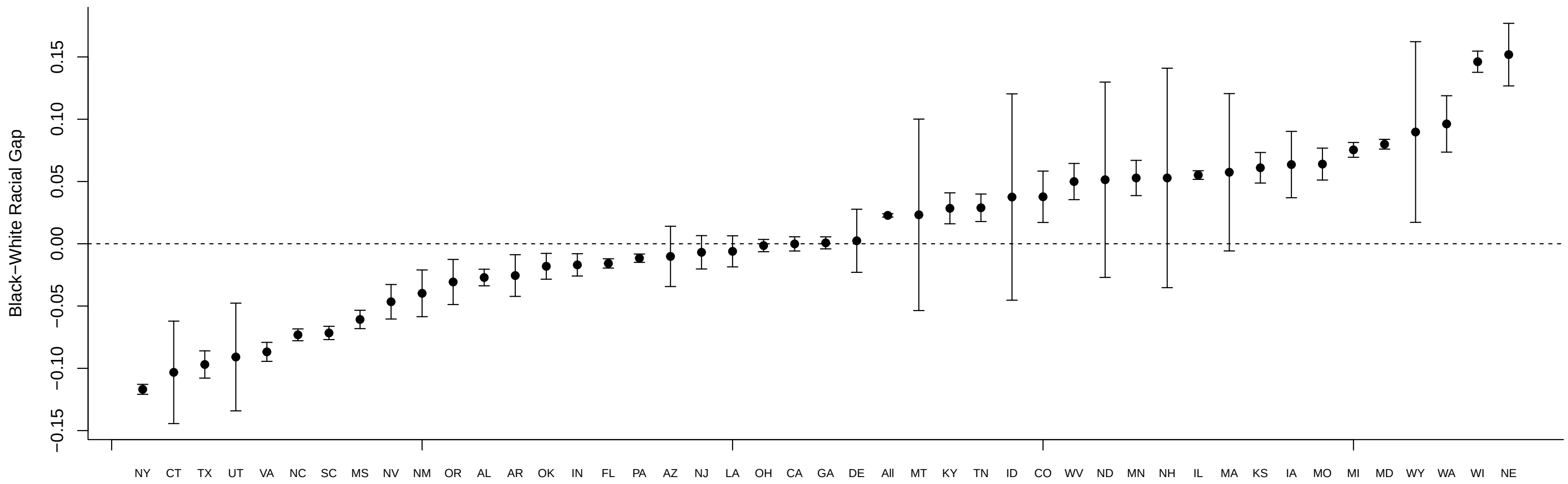
Notes: This table shows estimates of the structural model using Gibbs sampling. A burn-in period corresponding to the first half of the chain was used. The controls for other characteristics in columns 3 and 4 refers to in-group preferences by education, age, and dementia status, heterogeneity in demand for quality by these characteristics, and including these characteristics in nursing homes' admissions policies.

Figure A.1: Black/Non-Black Versus Hispanic/Non-Hispanic Segregation (Statewide Index of Dissimilarity)

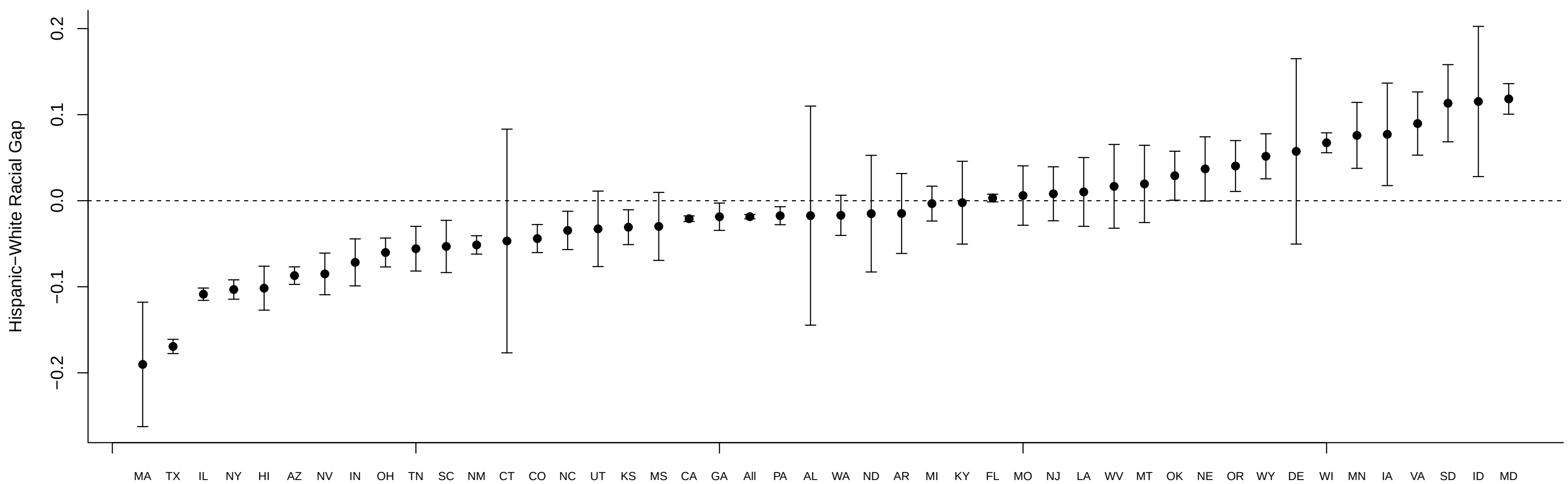


Notes: This figure shows a scatter plot of the dissimilarity index for black versus non-black residents and hispanic versus non-hispanic residents measured at the state level. Observations are weighted by the number of residents admitted to nursing homes in the state in question.

Figure A.2: Racial Gaps in Nursing Home Quality as Measured by LPN Staffing



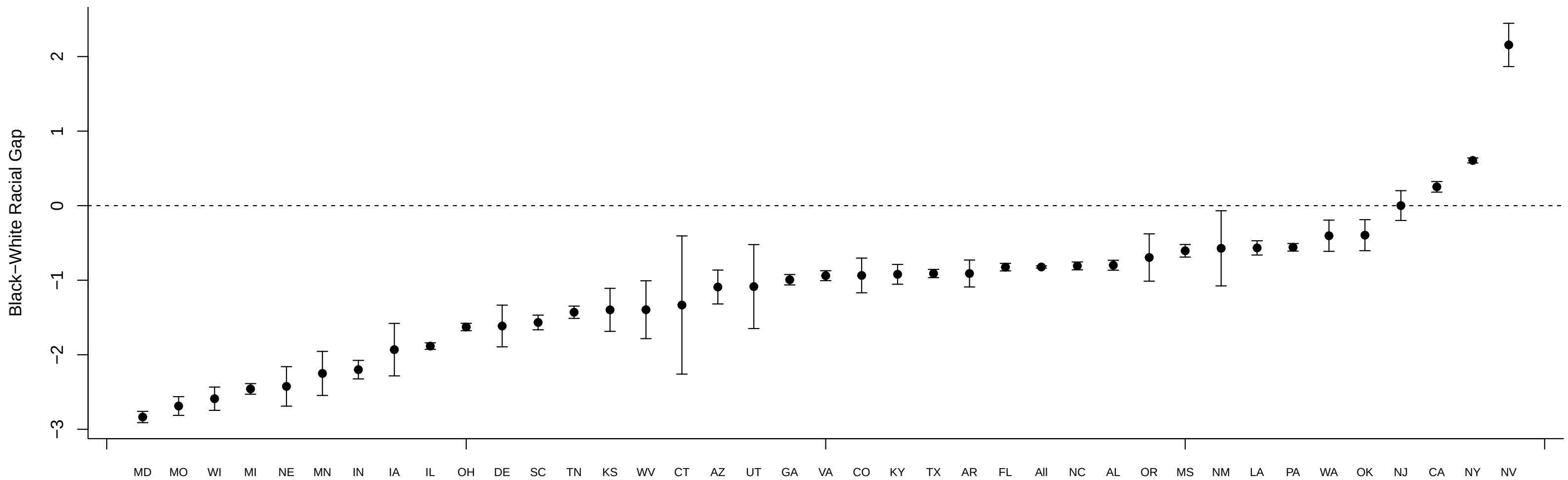
(a) Black-White Gap



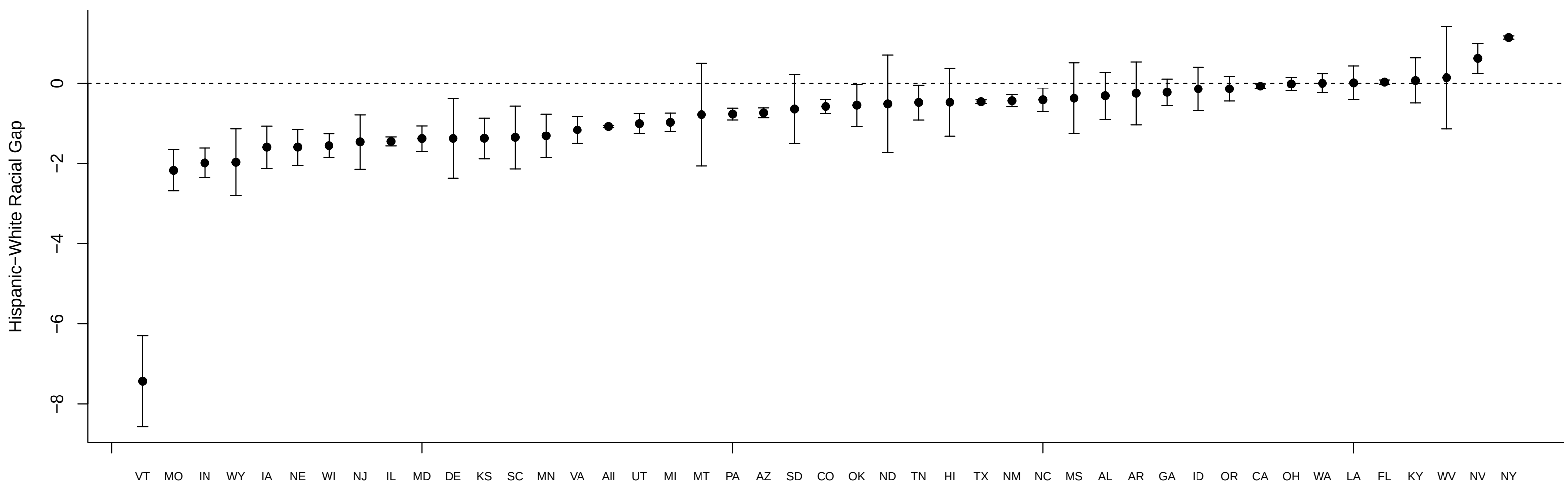
(b) Hispanic-White Gap

Notes: These figures display the estimated racial gaps in nursing home quality by state. Error bars indicate 95 percent confidence intervals for the estimates.

Figure A.3: Racial Gaps in Nursing Home Quality as Measured by Fewer Standard Deficiencies



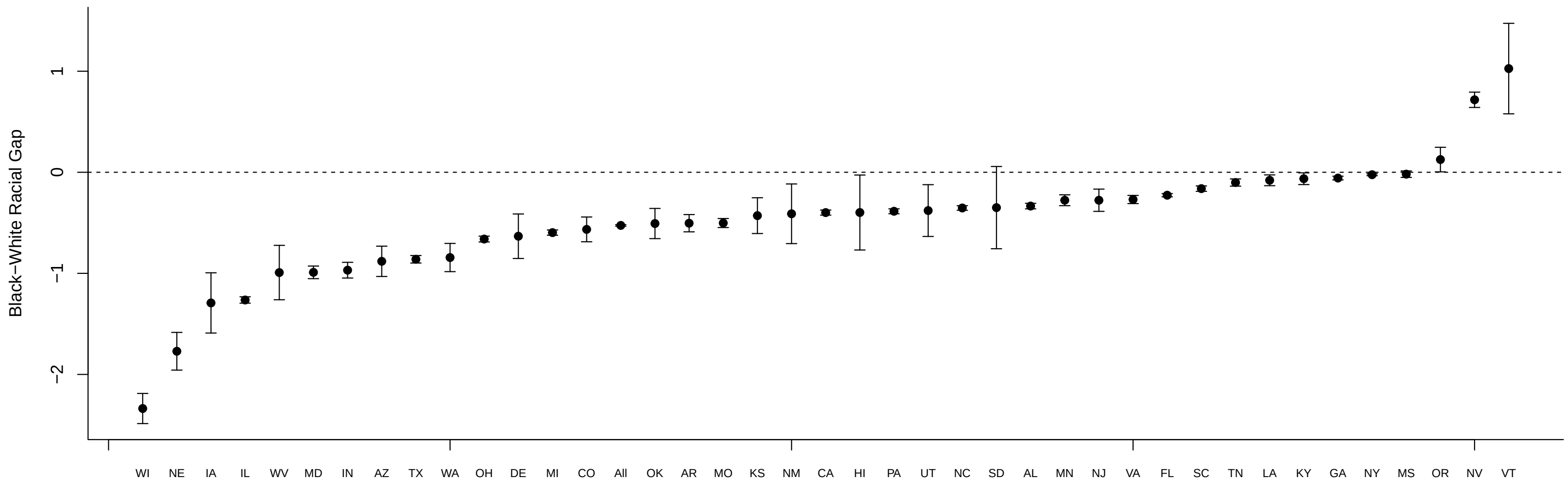
(a) Black-White Gap



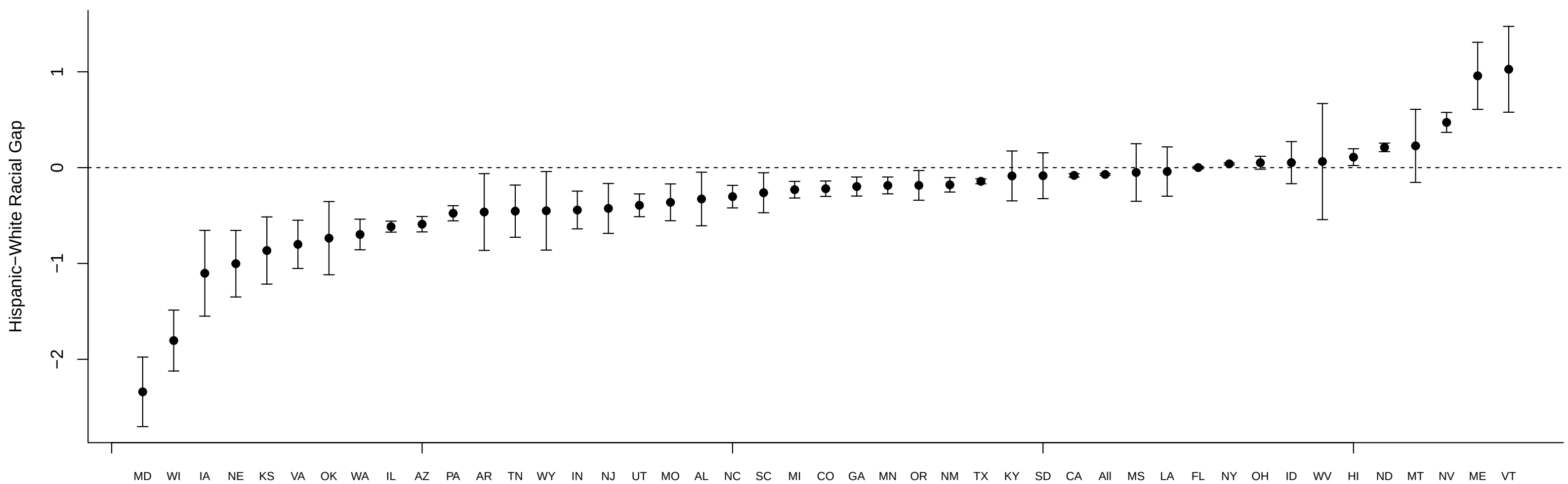
(b) Hispanic-White Gap

Notes: These figures display the estimated racial gaps in nursing home quality by state. Error bars indicate 95 percent confidence intervals for the estimates.

Figure A.4: Racial Gaps in Nursing Home Quality as Measured by Fewer Complaint Deficiencies



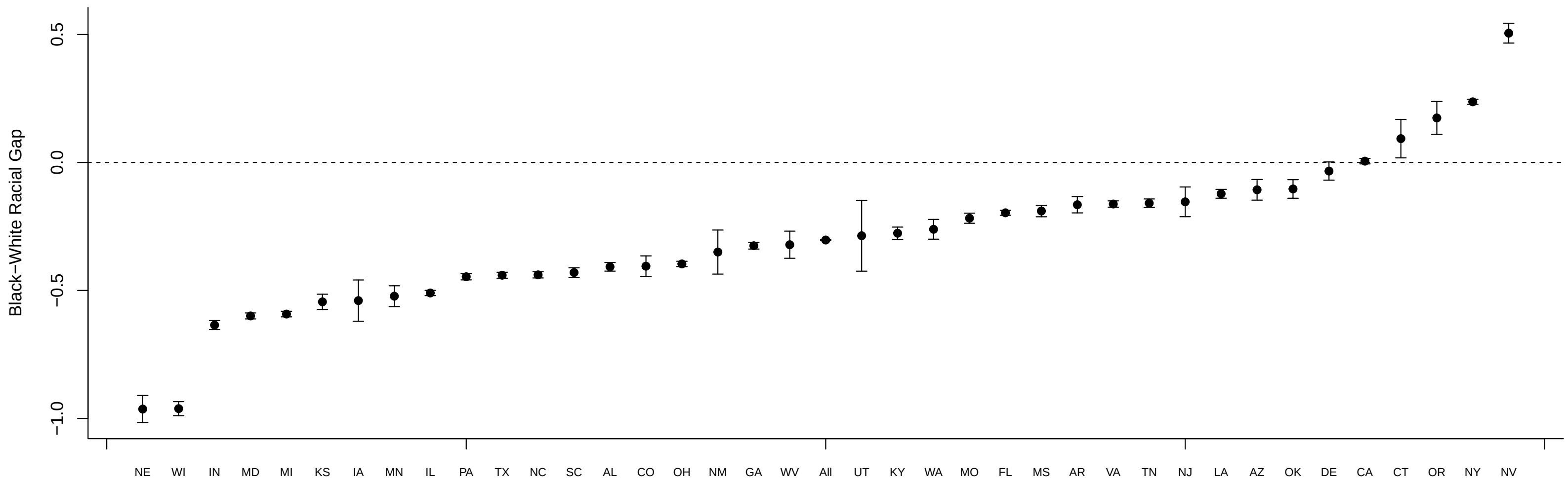
(a) Black-White Gap



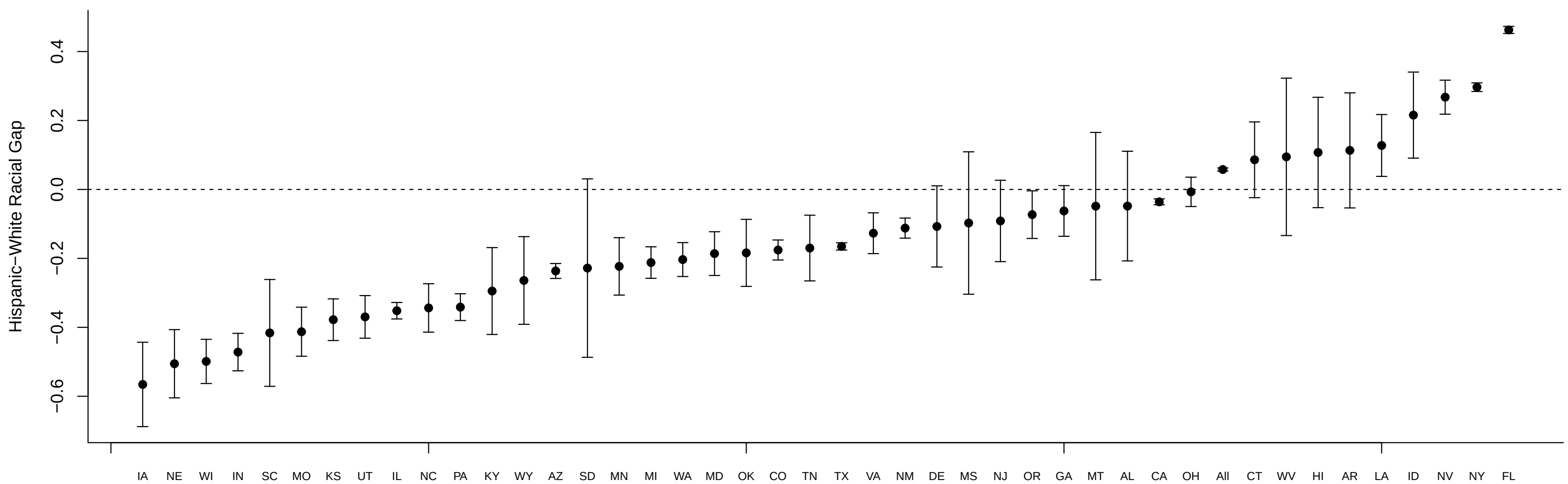
(b) Hispanic-White Gap

Notes: These figures display the estimated racial gaps in nursing home quality by state. Error bars indicate 95 percent confidence intervals for the estimates.

Figure A.5: Racial Gaps in Nursing Home Quality as Measured by Star Ratings



(a) Black-White Gap

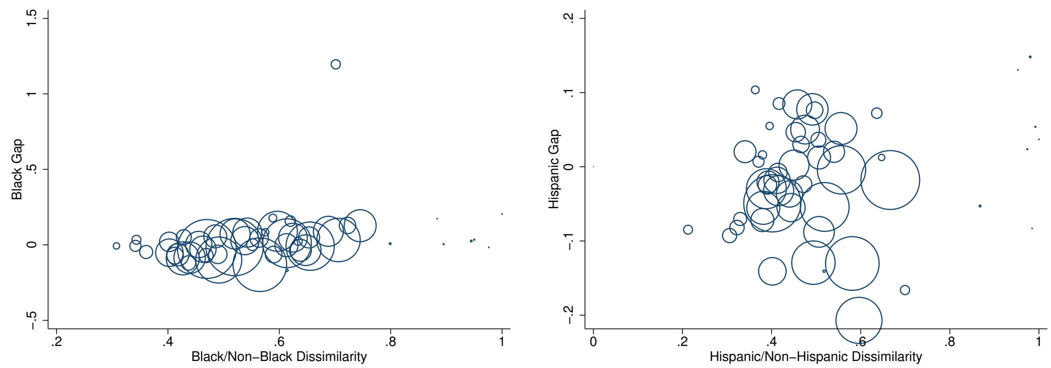


(b) Hispanic-White Gap

Notes: These figures display the estimated racial gaps in nursing home quality by state. Error bars indicate 95 percent confidence intervals for the estimates.

Figure A.6: Cross-Sectional Relationship Between State-Level Segregation and Disparities (LPN Staffing)

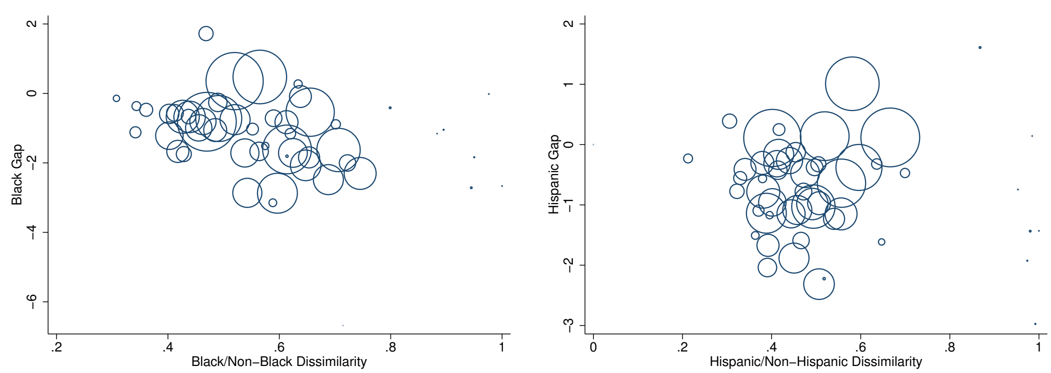
(a) Black/Non-Black (b) Hispanic/Non-Hispanic



Notes: These figures display scatter plots of the estimated racial gap (based on LPN staffing) against segregation at the state level. Observations are weighted by the number of residents admitted to the state.

Figure A.7: Cross-Sectional Relationship Between State-Level Segregation and Disparities (Fewer Standard Deficiencies)

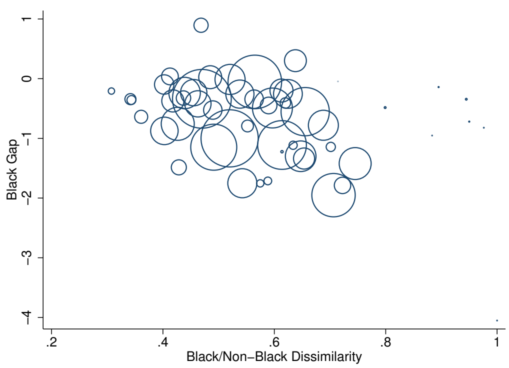
(a) Black/Non-Black (b) Hispanic/Non-Hispanic



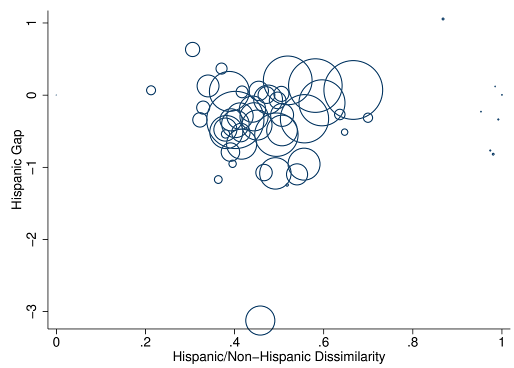
Notes: These figures display scatter plots of the estimated racial gap (based on fewer deficiencies) against segregation at the state level. Observations are weighted by the number of residents admitted to the state.

Figure A.8: Cross-Sectional Relationship Between State-Level Segregation and Disparities (Fewer Complaint Deficiencies)

(a) Black/Non-Black

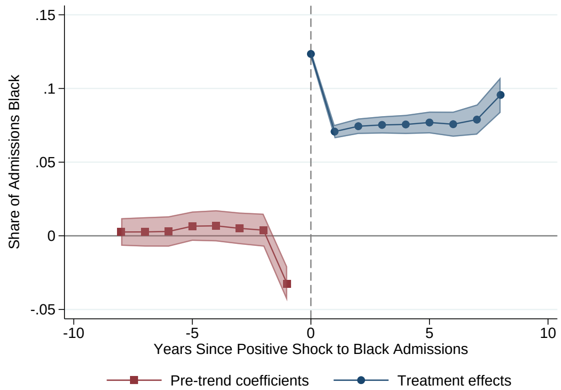


(b) Hispanic/Non-Hispanic

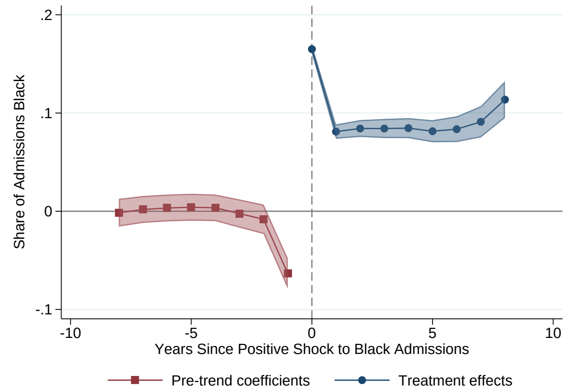


Notes: These figures display scatter plots of the estimated racial gap (based on fewer deficiencies) against segregation at the state level. Observations are weighted by the number of residents admitted to the state.

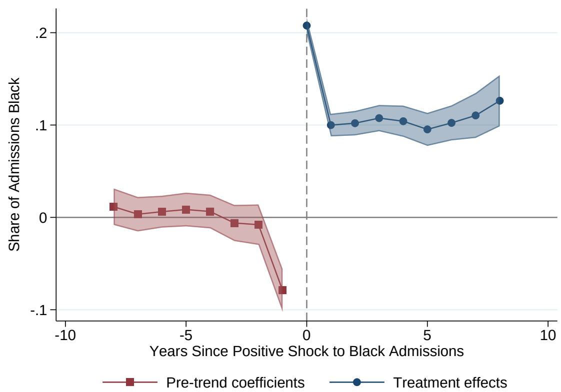
Figure A.9: Event Study on the Effect of a Positive Shock to the Share of Minority Admissions (Other Event Thresholds)



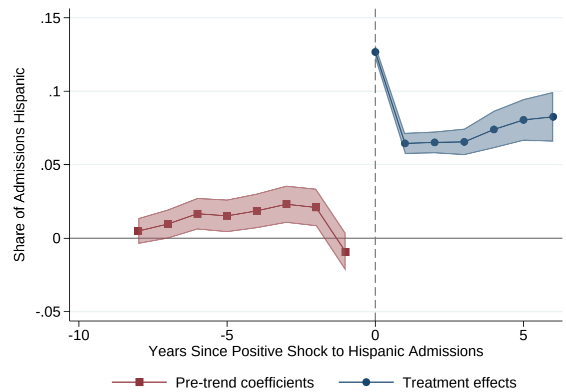
(a) >10pp. Increase in Share of Black Admissions



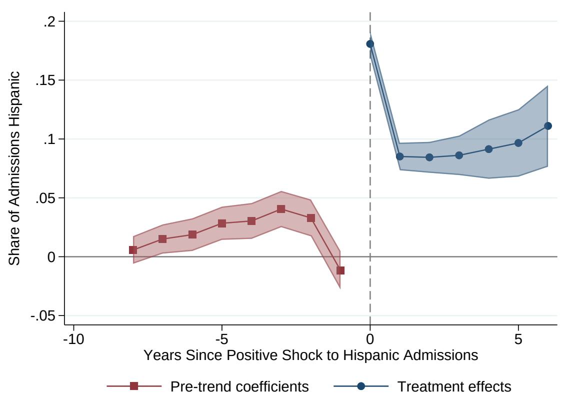
(b) >15pp. Increase in Share of Black Admissions



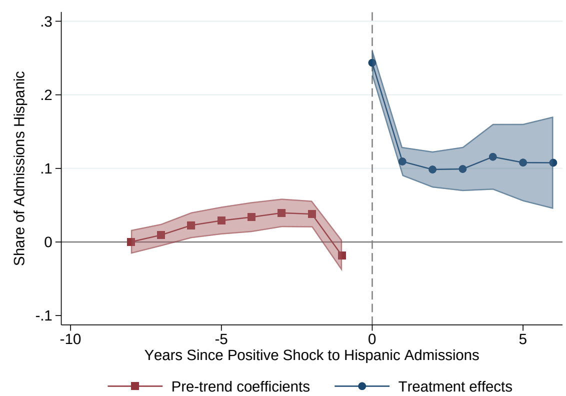
(c) >20pp. Increase in Share of Black Admissions



(d) >10pp. Increase in Share of Hispanic Admissions

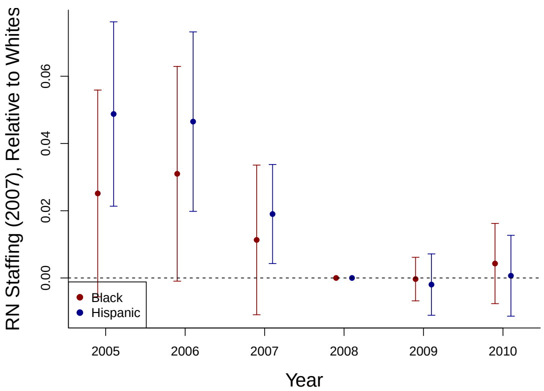


(e) >15pp. Increase in Share of Hispanic Admissions

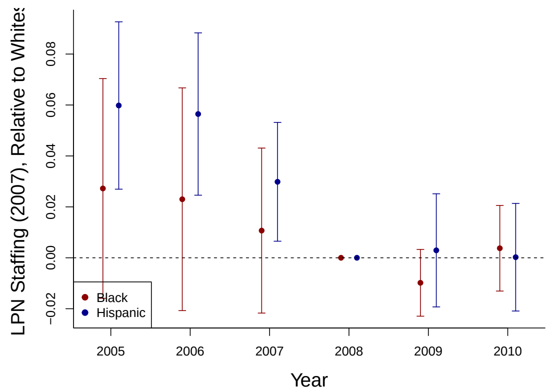


(f) >20pp. Increase in Share of Hispanic Admissions

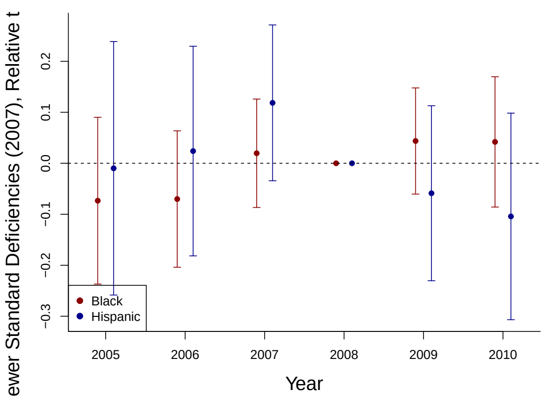
Figure A.10: Event Study on the Effect of Star Ratings on Nursing Home Choice



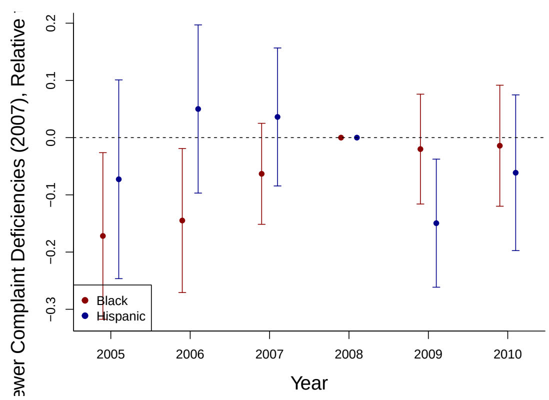
(a) RN Staffing (2007)



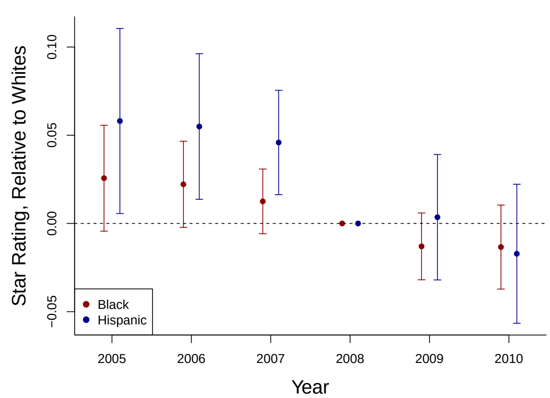
(b) LPN Staffing (2007)



(c) Fewer Standard Deficiencies (2007)



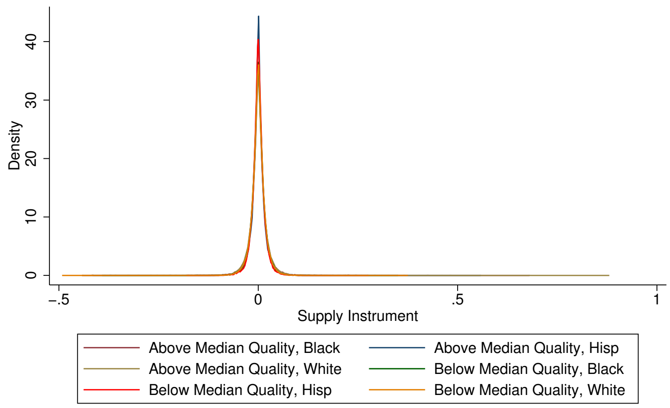
(d) Fewer Complaint Deficiencies (2007)



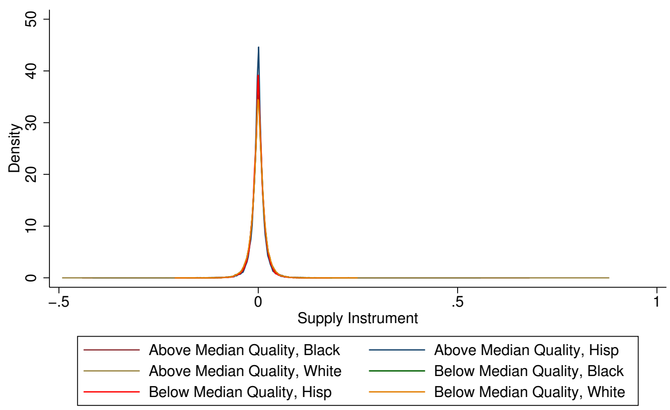
(e) Star Ratings (2009)

Figure A.11: Exclusion Restriction for Temporary Occupancy Fluctuations: Other Quality Measures

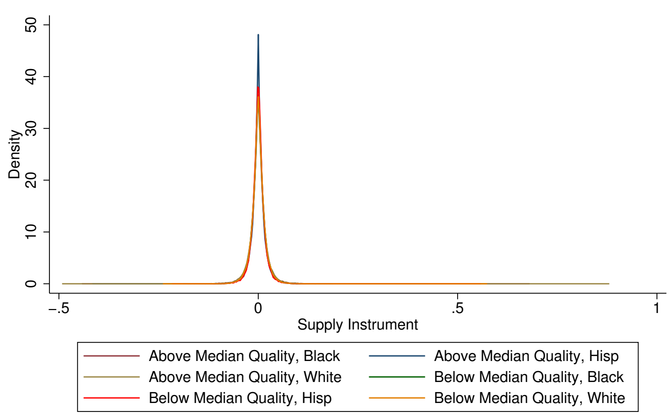
(a) LPN Staffing as Quality Measure



(b) Fewer Standard Deficiencies as Quality Measure

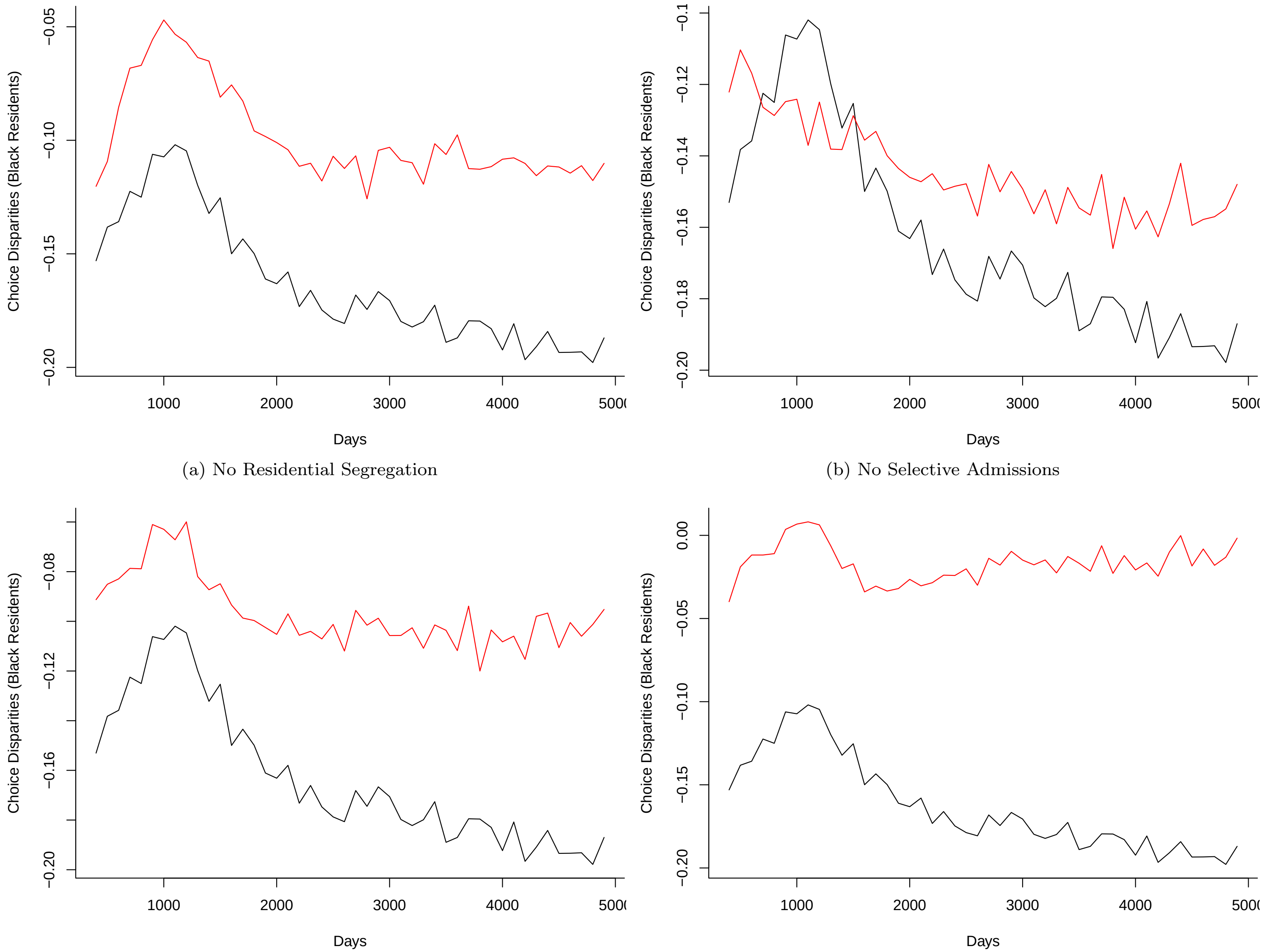


(c) Fewer Complaint Deficiencies as Quality Measure



Notes: These figures display kernel density plots of temporary occupancy fluctuations (defined as lagged 7-day log occupancy residualized of nursing home-month fixed effects) of nursing homes within 15 miles of each resident at their time of admission.

Figure A.12: Simulated Evolution of Disparities in Complaint Deficiencies Under Estimated and Counterfactual Parameters (Black Residents)



(a) No Residential Segregation

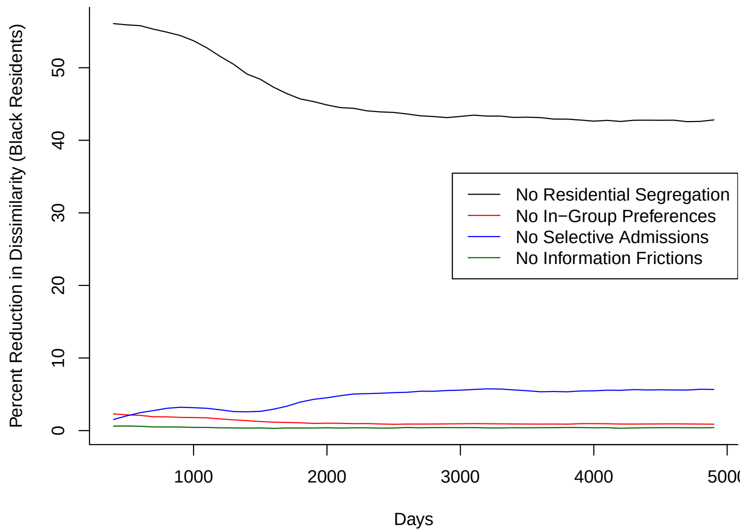
(b) No Selective Admissions

(c) No In-Group Preferences

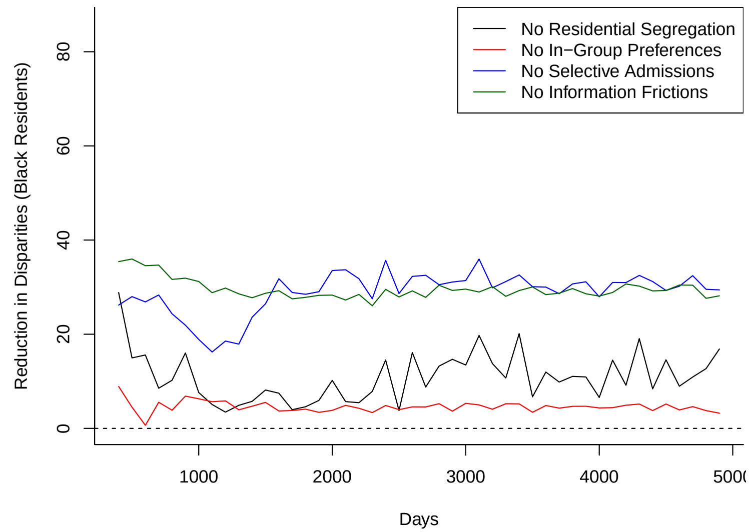
(d) No Information Frictions

Notes: These figures the simulated effect of various counterfactuals on racial segregation and disparities for residents in Floridian nursing homes over time.

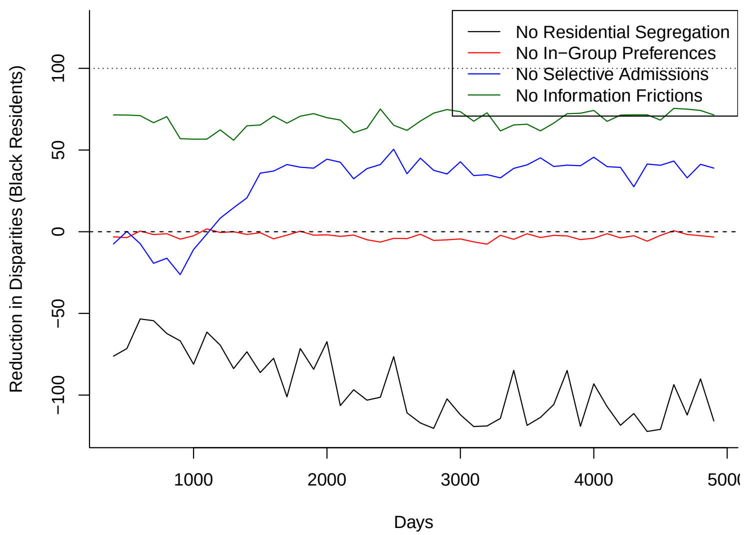
Figure A.13: Simulated Effect of Counterfactual Policies on Racial Segregation and Disparities (Black Residents)



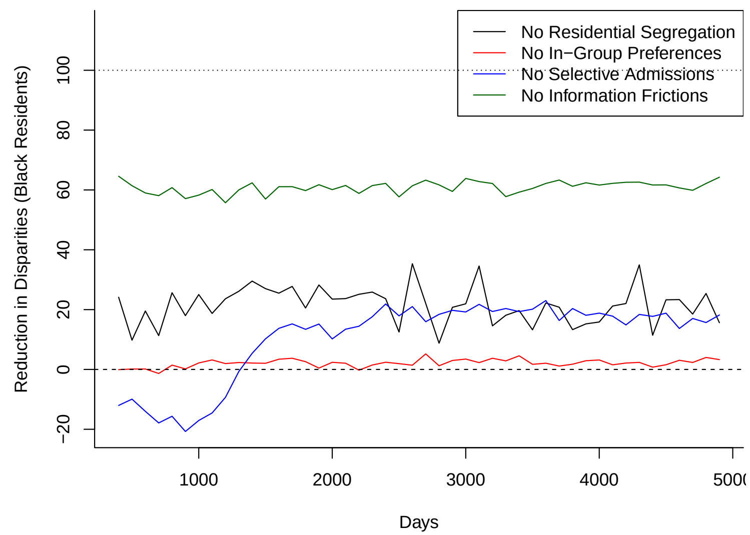
(a) Reduction in Segregation



(b) Reduction in Disparities (RN Staffing)



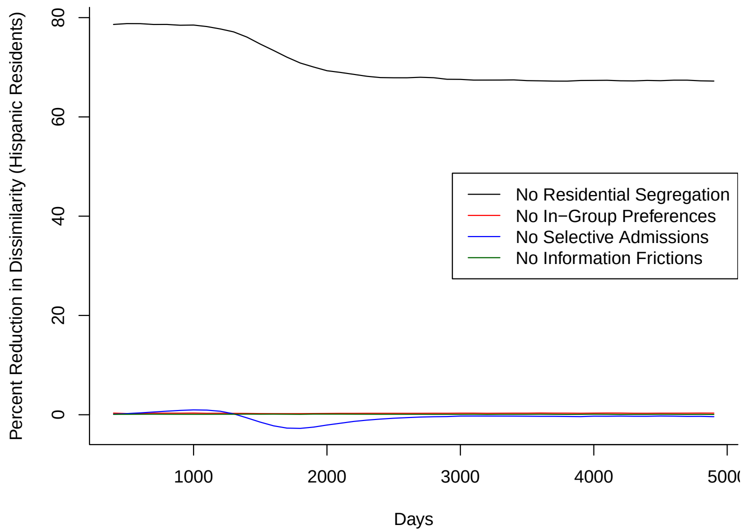
(c) Reduction in Disparities (LPN Staffing)



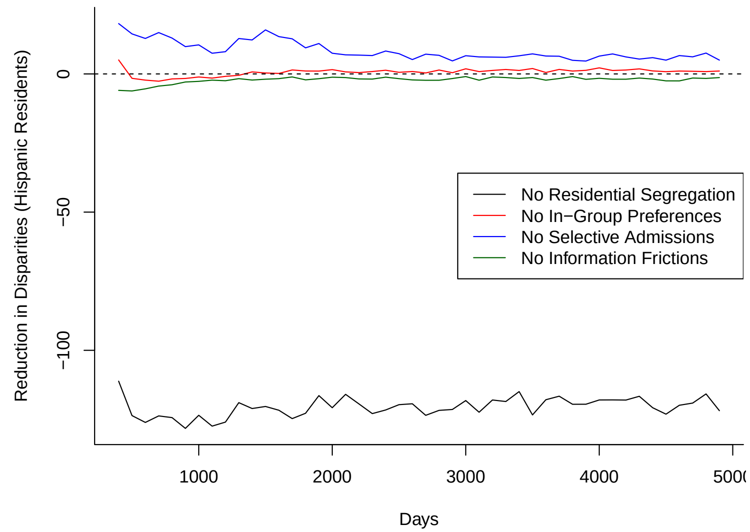
(d) Reduction in Disparities (Fewer Complaint Deficiencies)

Notes: These figures the simulated effect of various counterfactuals on racial segregation and disparities for residents in Floridian nursing homes over time.

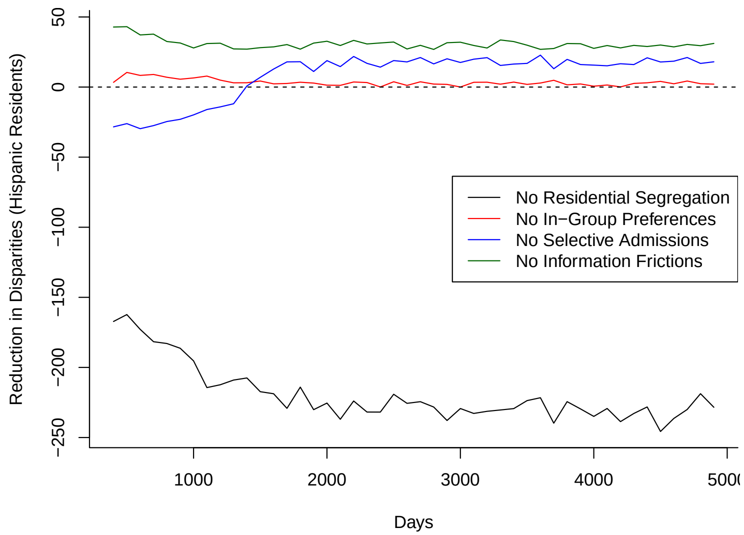
Figure A.14: Simulated Effect of Counterfactual Policies on Racial Segregation and Disparities (Hispanic Residents)



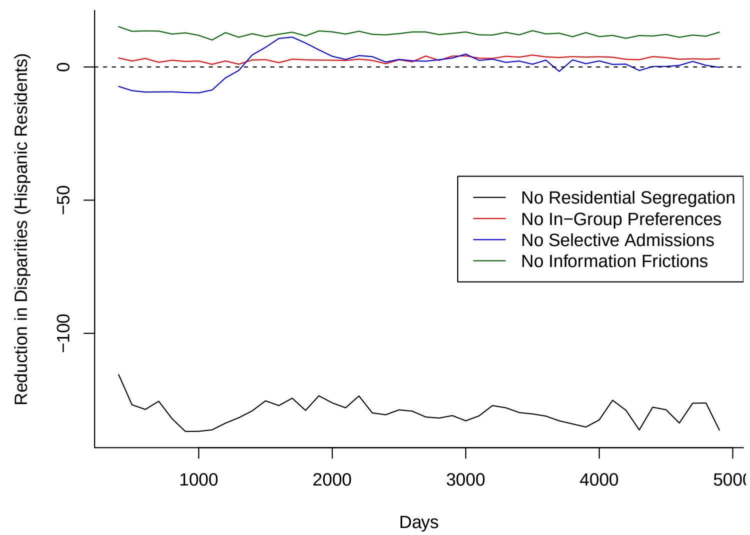
(a) Reduction in Segregation



(b) Reduction in Disparities (RN Staffing)



(c) Reduction in Disparities (LPN Staffing)



(d) Reduction in Disparities (Fewer Complaint Deficiencies)

Notes: These figures the simulated effect of various counterfactuals on racial segregation and disparities for residents in Floridian nursing homes over time. The black line shows the simulated evolution under the estimated parameters, whereas the red line shows simulated evolution under the counterfactual.