Demographic Differences in Social Exposure to High-Income People*

Victor Couture[†] Jonathan I. Dingel[‡] Allison Green[§]
Jessie Handbury[¶]

November 2023
Preliminary and Incomplete

Abstract

We study differences across demographic groups in their exposure to high-income individuals in shared spaces, using smartphone movement data. Black, Hispanic, and lower-income individuals are less exposed to high-income people. To distinguish preferences over the demographics of co-patrons from preferences for venue attributes and physical proximity to venues, we study choices of venues within business chains. We find remarkable regularities in social preferences across demographic groups: people prefer high-income and own-race co-patrons. Black and Hispanic individuals, however, face a trade-off between income and racial exposure. Within groups, individuals' preferences over co-patron demographics align with the demographics of their neighborhood of residence and can predict the neighborhood choices of movers.

Keywords: homophily, preference estimation, segregation, smartphone data
JEL codes: C55, D12, J1, L8, R2, R4

*We are grateful to Andrea Cerrato, Hongrui He, Shih-Hsuan Hsu, John Ruf, and Serena Xu for excellent research assistance. We thank Drew Breunig, Nicholas Shailas, Stephanie Smiley, Elizabeth Cutrone, and the team at PlaceIQ for data access and helpful conversations. The views expressed herein are those of the authors and do not necessarily reflect the views of PlaceIQ, NBER, nor CEPR. This material is based upon work supported by the National Science Foundation under Grant No. 2030056. This research was approved by the University of Chicago Institutional Review Board under protocol IRB20-0967. Dingel thanks the Cohen and Keenoy Faculty Research Fund at the University of Chicago Booth School of Business for support. Handbury thanks the Zell/Lurie Real Estate Center at Wharton and the Dean's Research Fund for financial support.

[†]University of British Columbia
[‡]University of Chicago Booth School of Business, NBER, and CEPR
[§]Princeton University
[¶]University of Pennsylvania and NBER

1 Introduction

American life is demographically fragmented: segregation by race and income are evident in domains from residential neighborhoods to media diets. Because social contacts are important for economic mobility, segregation of social interactions likely contributes to racial and income inequality. In this paper, we study differences across demographic groups in their exposure to high-income individuals. We measure exposure to high-income people in shared commercial spaces, estimate preferences over the racial and income composition of co-patrons, and use these estimates to quantify sources of cross-group differences in experienced income exposure.

To measure social exposure to high-income co-patrons, we use data on the movements of millions of smartphones in the United States in 2018 and 2019. We join smartphone data describing individuals' trips to residences and venues with building-level information on resident characteristics. With the combined data, we measure the socioeconomic composition of each venue's patrons and characterize eight groups' exposure to different co-patron mixes. The eight demographic groups are four racial-ethnic categories (Hispanic and non-Hispanic Asian, Black, and White) interacted with two income categories (split by median income).

We find large differences across groups in exposure to high-income co-patrons. Unsurprisingly, within each racial group, high-income individuals have greater high-income exposure. Within income groups, Black and Hispanic individuals have lower high-income exposure than White and Asian individuals. High-income Black individuals, for example, experience the same high-income exposure as low-income White individuals.

What explains these demographic differences in exposure to high-income individuals? They might reflect cross-group differences in access to venues, tastes in venue attributes, or preferences over co-patron demographics. Low-income individuals might prefer to associate with high-income co-patrons but live far from venues with high-income patrons or dislike the products offered at such venues. Or they might exhibit homophily: the tendency to associate with those similar to themselves.

To make progress on this question, we estimate preferences over co-patron demographics using variation across establishments within chain businesses. Because chains offer standardized services across multiple establishments, we can distinguish preferences over co-patron composition from tastes for service attributes. Our baseline estimation sample contains restaurants, the business category with the greatest number of chain establishments. The large number of observed choices, sometimes by the same individual, allow us to estimate preferences for social exposure in a way that one could not with data on home purchases or school enrollments.

We estimate social preferences over co-patron demographics using a model of venue choice. Our research design exploits the trade-off between the cost of a longer trip and the benefit of a preferred co-patron composition, yielding estimates of social preferences expressed in terms of willingness to travel. We specify preferences as a flexible function of the share of own-race co-patrons and the share of high-income co-patrons. This flexibility allows us to capture interactions between race and income composition and to distinguish pure homophily from aversion to being an extreme minority or a preference for integrated demographics.

We find remarkable regularities across demographic groups in their preferences over co-patron composition. Rich and poor individuals exhibit similar levels of racial homophily. White, Black, and Hispanic individuals have similar levels of racial homophily (with that of Asian individuals being somewhat stronger). All groups display some preference for high-income co-patrons, but those preferences are only monotonic for high-income individuals. Low-income individuals prefer establishments with a mix of low- and high-income co-patrons. The willingness to travel for high-income exposure is broadly similar across racial groups, but smallest among White individuals. These preferences for social exposure are economically large. Individuals are willing to travel two to three additional kilometers to visit a venue in the top decile of either the own-race or high-income distribution rather than a bottom-decile venue.[1] This translates into willingness to pay of a few thousand dollars per year, close in magnitude to willingness to pay for schools with high test scores (e.g., Black, 1999).

Given these regularities in social preferences, why is experienced social exposure so different across demographic groups? In our framework, four factors determine exposure to high-income co-patrons. First, individuals live in different cities, which differ in their patron demographics. Second, individuals may patronize different chains because of their different attributes (including co-patron composition). Third, travel is costly, and co-patron composition varies across space. Finally, individuals have social preferences over the race and income composition of their co-patrons within venues. The first two factors are small: metropolitan area of residence and choice of restaurant chain explain little of the variation in social exposure to high-income co-patrons. Our estimated model suggests that cross-group differences in exposure largely reflect the roles of social preferences and neighborhood of residence.

Racial differences in high-income exposure are generated by preferences combined with the joint distribution of income and race across venues, and in the population as a whole. Preferences over co-patron income are similar across racial groups, but the nature of the joint

[1] The exception is low-income individuals' preference for income exposure. Low-income individuals have weaker income preferences and prefer integrated venues in the income dimension, with a share of high-income co-patrons near the median.

distribution of income and race combines with racial homophily to generate differences in exposure to high-income co-patrons. For instance, high-income Black and White individuals share similar willingness to travel to high-income venues. Black individuals, however, visit venues with much smaller shares of high-income co-patrons, even conditional on the cost of distance from their residential locations to those venues. This reflects racial homophily. Because heavily Black (or Hispanic) venues generally have lower-income co-patrons, Black (or Hispanic) individuals face a trade-off between visiting heavily high-income venues and visiting heavily own-race venues that White and Asian individuals do not face.

The gap in income exposure between low- and high-income individuals within racial and ethnic groups reflects gaps in preferences and residential sorting. For example, high-income White individuals choose to live in neighborhoods near venues with many high-income and many White co-patrons, and conditional on their proximity to venues with different co-patron mixes, they are also more likely to choose venues with more high-income and White co-patrons. Income exposure is, in fact, often over-determined: it can be explained either by social preferences or by neighborhood choice.

To further examine the link between social preferences and neighborhood choice, we estimate how preferences for social exposure vary across individuals in the same demographic group but residing in neighborhoods with different demographic mixes. We find that groups live in neighborhoods that match their preferences for social exposure: individuals living in higher-income neighborhoods have stronger income preferences, and individuals living in more heavily own-race neighborhoods have stronger racial preferences. This alignment of individuals' social preferences and the dominant demographics of their residential neighborhoods suggests that social preferences might, in addition to determining venue choice, be a determinant of neighborhood choice.

Our model is agnostic on how individuals choose residential neighborhoods, but we can test for such sorting patterns with a movers design. Specifically, we estimate the preferences of individuals who move between metropolitan areas transitioning between pairs of neighborhoods with different demographic mixes. The estimated preferences are consistent with people sorting into neighborhoods based on their social preferences. Social preferences predict neighborhood choice in the sense that individuals move to neighborhoods with demographics that align with their social preferences estimated from pre-move visits. For example, an individual exhibiting weaker racial homophily is more likely to move to a more integrated neighborhood than their counterpart exhibiting stronger racial homophily.[2] Consistent with contact theory, preferences for the local demographic mix slightly strengthen after the move. This mover analysis also validates our model specification: the estimated preferences do not

[2] This pattern is consistent with social connections predicting residential moves (Büchel et al., 2020).

shift discontinuously when an individual's choice set changes. While this investigation of movers' preferences is limited by sample sizes and a short time horizon, it demonstrates the potential for mobility data to advance our understanding of social preferences.

This paper contributes to a literature documenting segregation in non-residential domains. Economists have documented racial segregation of friendship networks (Echenique and Fryer, 2007), gender segregation of retail venues (Caetano and Maheshri, 2019), and income segregation of universities (Chetty et al., 2020). Closer to this paper is recent work documenting segregation in the places people visit by race (Davis et al., 2019; Athey et al., 2021; Baldenius et al., 2023), socioeconomic status (Moro et al., 2021; Xu et al., 2019; Abbiasov et al., 2022; Cook, 2023; Massenkoff and Wilmers, 2023), or student status (Cook, Currier, and Glaeser, 2022). We document experienced segregation within commercial spaces by income and race jointly. In this vein, Wang et al. (2018) show that residents of Black and Hispanic neighborhoods travel similar distances as others but visit high-income neighborhoods less. We use the trade-off between travel distance and demographic composition to identify individuals' preferences over co-patron demographics. We find that Black and Hispanic individuals would prefer both high-income and own-race co-patrons, but few such venues exist.

Much of the literature on preferences over social environments has focused on residential location decisions (e.g., Schelling 1971; Card, Mas, and Rothstein 2008). These studies aim to distinguish preferences over neighbors' income and racial demographics from preferences over other attributes such as neighborhood amenities (Caetano and Maheshri, 2021; Bayer et al., 2022; Davis, Gregory, and Hartley, 2023; Schönholzer, 2023). We complement this work by estimating preferences for the demographics of fellow customers within business chains, a setting with more uniform attributes and many more observed choices.[3]

Finally, our paper complements growing evidence on the economic benefit of social connections to higher-income people. Social connections help workers find jobs through referrals (Bayer, Ross, and Topa, 2008; Barwick et al., 2023). One's number of high-socioeconomic-status Facebook friends is a robust correlate of social mobility (Chetty et al., 2022*a*,*b*). Our mobility data, however, measure social exposure, not social connections. Atkin, Chen, and Popov (2022) use smartphone data to show that serendipitous encounters in Silicon Valley in the kind of venues we study produce more patent citations between the connected employers.

[3]In the housing market, racial differences in socioeconomic status mean that racial minorities face a trade-off between sorting into high own-race and high-income neighborhoods (Sethi and Somanathan, 2004; Bayer, Fang, and McMillan, 2014; Reardon, Fox, and Townsend, 2015). We show that minorities face the same trade-off when choosing social spaces throughout their daily lives. Like some studies of residential decisions (Bayer and McMillan, 2005; Aliprantis, Carroll, and Young, 2022), we find that racial homophily plays a dominant role in this setting.

Beyond commercial gains, Anderson (2011) argues that overlapping visits to shared spaces by people of different backgrounds may be a basis for building understanding and tolerance.

2 Data

To measure social exposure and estimate social preferences, we need information on the home location and demographic characteristics of a large sample of individuals, and on the venues that they visit. This section describes the construction of our estimation sample from two main data sources: smartphone movement data on visits to chain restaurant venues, and building-level data on demographic characteristics. Appendix A offers more details on each data source.

2.1 Data sources

Our smartphone movement data is from PlaceIQ, a location data and analytics firm. PlaceIQ aggregates pings from smartphone applications that request locational services from the devices' operating system.[4] Pings originating from different applications on the same smartphone are linked to a unique advertising identifier, which we denote a "device." These pings are intersected with a two-dimensional map of polygons corresponding with buildings or outdoor locations such as public parks, which we denote "venues." A spatial and temporal cluster of pings by a given device in or close to a venue constitutes a "visit" to that venue. PlaceIQ uses the timing of the first and last ping in the visit ping cluster to compute a lower bound estimate for visit duration.

Information on the demographic characteristics of each device comes from building-level data that include the income bracket, race, age bracket, gender, and education of individuals living at an address. PlaceIQ does not disclose the third-party provider of this data, so we discuss the reliability of its demographic information later in this section. These demographic data are aggregated across all units within a building. Thus, for single-family houses we observe the demographics of the household, while for multi-unit buildings we observe building-level averages. We impute demographics to devices using their inferred residence, which is the residential building where the device regularly spends time at night.

[4]Some applications collect location data only when in active use, while others collect location data while running in the background. We do not know the set of applications contributing data.

2.2 Estimation sample

In this subsection, we describe the selection of devices, venues, and visits in our estimation samples. For the purpose of estimating social preferences, we create a restricted sample of devices and visits for which we know demographic information and trip purpose with higher confidence. For the purpose of measuring the demographic composition of each venue, we use a sample of devices and visits that is as broad as possible. Our sample covers the 100 largest metropolitan areas from June 1, 2018 to December 31, 2019

Device selection criteria Around 66 million devices in our smartphone sample have exactly one home assignment over our 18 month sample period.[5] Around 46 million of these devices live in buildings for which we have demographic data. We classify building-level demographics in terms of two income groups and four racial/ethnic groups: the share of a building's residents with household income above $75,000 (the bracket cutoff closest to the national median in 2019) and the shares of a building's residents who are Black, White, Asian, and Hispanic. We use visits by all devices for which we have building-level data to measure the demographic mix of co-patrons in each venue, applying device demographics probabilistically. To identify social preferences, we limit our estimation sample to choices by devices whose demographic characteristics can be reasonably approximated with a single profile. Specifically, we restrict our estimation sample to 36 million devices that live in buildings with relatively homogeneous characteristics, where at least two-thirds of the residents belong to the same income and racial/ethnic group. 93% of buildings are racially homogeneous and 99% of buildings are income homogeneous, consistent with most of these buildings being single-family homes, in which a large majority of Americans live. The high share of homogeneous buildings underscores a key advantage of using building-level data instead of Census tables (as in, for instance, Athey et al. 2021) to identify the demographic characteristics of devices.

Venue selection criteria Given our within-chain identification strategy, we only consider large chains. Our baseline estimation focuses on restaurants, which have by far the largest number of chains, establishments, and visits. We also characterize co-patron exposure in banks, big box retail stores, convenience store/gas stations, grocery stores, gyms, and pharmacies.

Table A.3 compares the number of venues we observe in the 10 largest restaurant chains to counts from external sources. We observe on average 87% of venues within these chains,

[5] Around 10% of devices move during our sampling period. We drop these from our baseline estimation sample. We return to studying these movers in Section C.2.

with a low of 69% for Starbucks. Since the smallest spatial unit of observation in our data is a building, we exclude venues that contain multiple establishments, such as shopping malls. To avoid concerns over store entry and exit, we only keep venues with at least one visit prior to the beginning of our estimation window (June 1, 2018) and one visit after the end of our estimation window (December 31, 2019). This excludes around 10% of chain venues from estimation sample.

Visit selection criteria To measure the average demographic composition of co-patrons in a given venue, we use all visits to that venue by devices for which we have demographic data over our 18-month study period. To identify social preferences, however, we restrict attention to direct trips to a venue that originate and end at home.[6] Considering only round trips from home ensures that a trip's only purpose is visiting a venue. This selection eliminates confounding factors due to trip chaining, and allows us to identify preferences within a standard venue choice model like that we propose in Section 4. We also exclude visits with duration longer than three hours, as these are likely from venue employees.[7] Overall, the sample of restaurant visits we use to estimate social preferences includes more than 14 million direct trips to more than 27,000 restaurant chain venues by almost 4 million devices who live in homogeneous buildings.

2.3 Data quality and representativeness

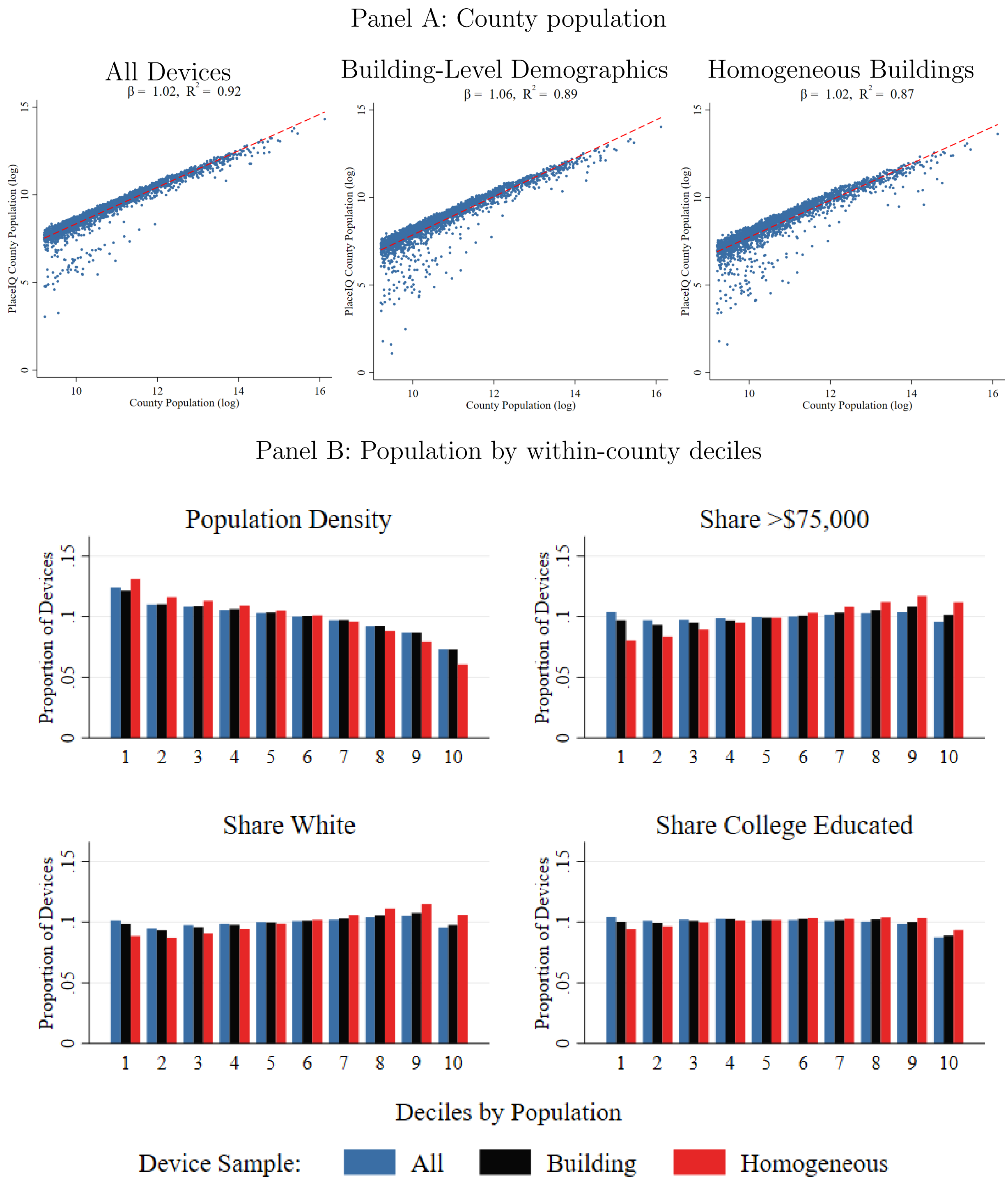
In this subsection, we first assess whether our device selection criteria biases our estimation sample. We then evaluate the reliability of the demographic information in the building level data.

Couture et al. (2021) show that devices active in the smartphone data are broadly representative spatially, and they make visits that resemble what travelers self-report in the National Household Travel Survey. Figure 1 shows that the additional selection criteria we impose on our estimation sample generate only limited spatial biases. Panel A plots the log number of devices residing in a county against the 2019 Census population estimates for three device samples. The "All Devices" sample includes all devices that have exactly one

[6] We define a direct visit as a visit to a venue where the immediately preceding and succeeding visits were to a device's home and within an "activity day." An activity day is a calendar day offset by three hours such that it begins and ends at 3 AM. We adjust from a calendar to an activity day to reflect late-night trips to venues. Davis et al. (2019) and Miyauchi, Nakajima, and Redding (2021) study consumption trips that can originate at the workplace.

[7] Note that visit duration is measured with error, but can be reasonably interpreted as a lower bound for actual duration. A visit is registered when a smartphone application collects a ping in a venue, not when they first enter the venue, so a device may spend more time at a venue than we observe.

Figure 1: Comparing estimation sample to full smartphone data



NOTES: Panel A compares the number of devices residing in a county (vertical axis) to the Census's estimated 2019 residential populations using three different device selection criteria: (1) all devices residing in exactly one residential building between June 1, 2018 and December 31, 2019; (2) among those devices in (1), the devices whose building-of-residence has demographic data available; (3) devices whose building-of-residence is comprised of at least 67% one income group and racial/ethnic group. We exclude counties with a Census population of less than 10,000 people. Panel B depicts the share of devices living in block groups within each within-county population decile for four characteristics: population density, population share of high-income ($> \$100,000$) residents, population share of white residents, and population share of residents who have obtained a bachelor's degree. Panel B depicts these decile shares for the three populations of devices depicted in Panel A.

home assignment. The "Building" sample only includes devices with building-level demographic data. The "Homogeneous Building" sample only includes devices living in mostly homogeneous buildings (> 67% one group in both race/ethnicity and income.) These regressions of device sample size on Census county population yield an R^2 of at least 0.87 for each of the three device samples. This suggests that our sample is representative of county-level population even restricting attention to devices living in homogeneous buildings.

Panel B of Figure 1 further tests for within-county spatial biases. The figure shows the share of devices living in within-county population deciles along four different demographic dimensions: population density, share high-income, share White, and share college educated.[8] If device samples were drawn in exact proportion to actual populations as measured in the Census, each bar would be of equal height (0.10). We show these results for the three different samples above. The bar heights are very similar in the "All" and "Building" device samples. This alleviates concerns over spatial bias in the building-level data that we use to compute the demographic composition of venues. When we restrict the sample to homogeneous buildings, we see a more substantial bias away from the highest density block groups, with about six percent of devices living within the top density decile,and a slight bias towards more heavily white and high-income block groups. So our estimation sample of devices living in homogeneous buildings is broadly spatially representative, with the exception of devices living in the top density decile (i.e., in multi-unit buildings) being somewhat underrepresented.

Finally, we evaluate the reliability of the demographic information in the building-level data. Here, we summarize the investigation that we conduct in Appendix A. We first show that although the devices for which we have building-level data are drawn across block groups in a spatially representative way (Figure 1), our"Building" sample still contains more White and high-income devices than Census tables. However, the cross-county correlation between the share of devices within a given demographic group in the Census and that share in the building-level data remains above 0.8 for all demographic groups. These deviations from perfect representativeness are expected in smartphone samples, but they warrant some caution when measuring the demographic composition of co-patrons within restaurant venues. We therefore follow Cook, Currier, and Glaeser (2022) and report results that highlight differences in social exposure across demographic groups, instead of absolute levels that may overstate exposure to high-income devices.

To validate the demographic information in the building level data, we show that it reli-

[8] These deciles are computed using within-county variation in 2015-2019 ACS block group information. Using national population deciles, we see a more pronounced bias towards whiter, less densely populated block groups. See Couture et al. (2021).

ably predicts differences in behavior between residents of neighboring houses. For instance, we find that residents of high-income buildings are more likely than their low-income neighbors to visit chains preferred by high-income people, such as Starbucks. We proceed as follows: First we determine the chains preferred by high income individuals, by assigning an income level to each device based on its neighborhood (block group) of residence, and ranking restaurant chains by how likely high-income people are to visit them relative to low-income people.[9] We then compare this restaurant chain ranking, obtained using only demographic information available in the Census, with an alternative ranking obtained using only demographic information from the building-level data. So we obtain an alternative ranking of chains preferred by high-income individuals, based on visit propensity of devices living in high- and low-income buildings within the same neighborhood. This chain ranking derived from building-level demographic information has a rank correlation of 0.8 with the analogous ranking derived from Census demographic information. We replicate this comparison for racial instead of income groups, and for convenience stores/gas station (the second largest establishment category) instead of restaurants, and find similarly high correlations. We draw two conclusions from this exercise, which we leverage in our empirical analysis. First, behavior predicted from the building-level demographic data is consistent with behavior predicted using Census demographic data. Second, the building-level data provides information not available in the Census, because it allows us to predict behavior within a neighborhood.

3 Patterns of social exposure

This section documents how exposure to different types of co-patrons varies by demographic group. We first show, for each demographic group, the full distribution of visits to chain restaurant venues by racial and income mix. We then show how these differences in visit patterns translate into disparities in income exposure across groups, and how these exposure disparities remain stable across a broad spectrum of venue types.

Figure 2 shows the racial and income composition of venues, and the propensity of each group in our estimation sample to visit venues by co-patron demographics. Each dot in the plot represents a restaurant within the 100 largest CBSAs. We compute visit propensity using a non-parametric kernel regression of a demographic group's share of visits to a venue on co-patron characteristics. Color variation represents variation in visit propensity: blue

[9] We only compute this ranking by comparing block groups within the same census tract, to avoid biases in our preference ranking due to differences in venue choice set. Such biases would arise if travel is costly and some chains co-locate with high-income people.

venues are more likely to be visited than average and darker shades of blue represent higher probabilities.

These plots document important regularities in visit propensity across demographic groups. For all groups, visit propensity is increasing in own-race share. Within each race, higher-income individuals visit venues with greater shares of high-income co-patrons than their low-income peers. These visit patterns echo familiar patterns of residential segregation by race and assortative matching by income. However, the distribution and availability of venues varies starkly across racial groups. Unsurprisingly, many more venues are predominately White than Black, Hispanic, and Asian. Heavily white venues vary significantly in their income composition, whereas heavily Black and Hispanic venues tend to be predominantly low-income. Finally, Asian individuals face very few venues with high share of Asian co-patrons, regardless of co-patron income.

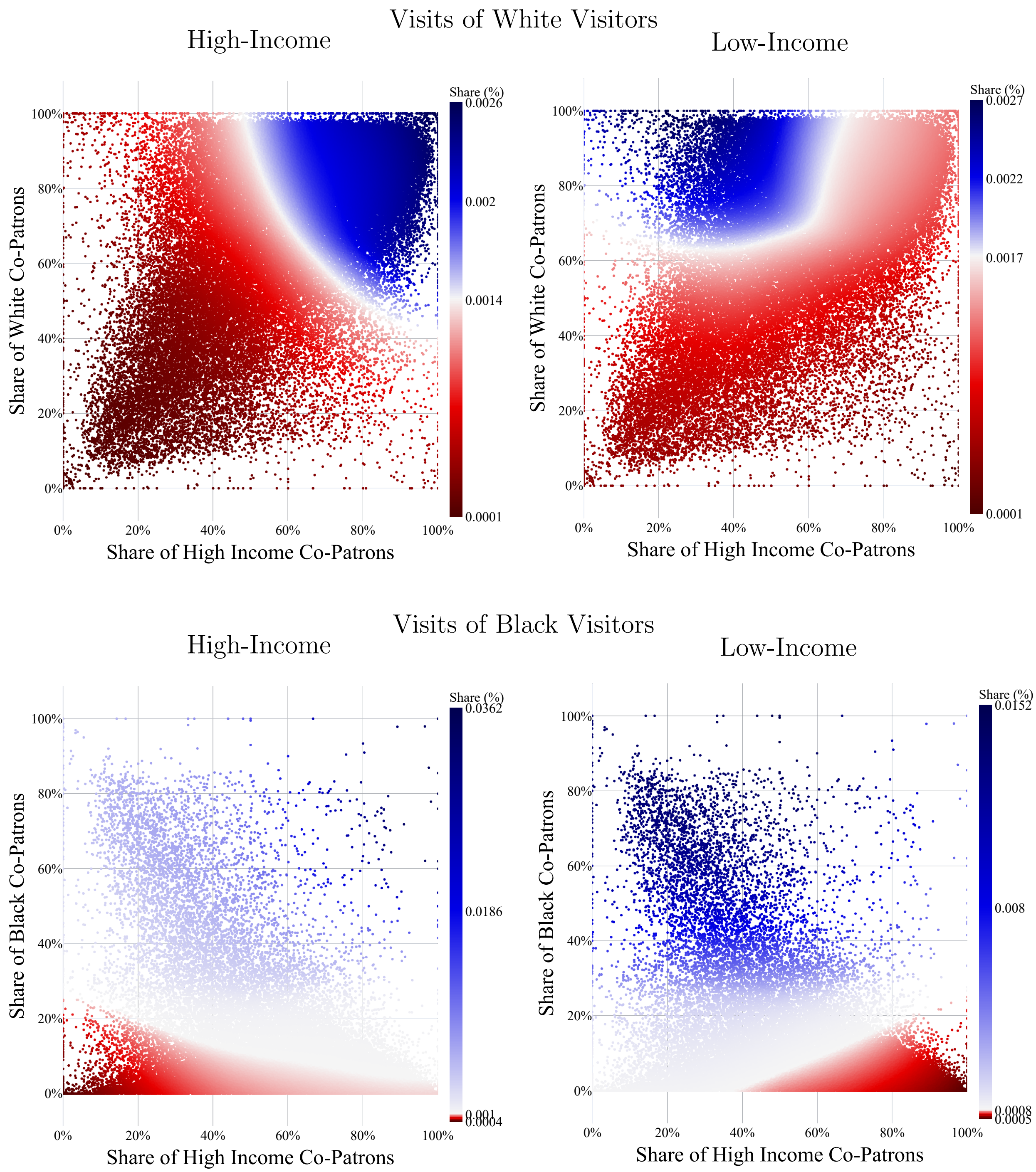
Table 1: Exposure to High-Income Co-Patrons

		Low Income				High Income			
	Uniform	White	Black	Hispanic	Asian	White	Black	Hispanic	Asian
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
HVH Visits to restaurant chains	0.60	-0.02	-0.16	-0.15	-0.06	0.14	-0.00	0.05	0.14
All Visits to restaurant chains	0.60	-0.02	-0.13	-0.13	-0.04	0.12	-0.00	0.04	0.12
All Visits to all chains	0.60	-0.03	-0.16	-0.14	-0.05	0.11	-0.01	0.04	0.12
All Visits to all non-residential establishments	0.60	-0.03	-0.15	-0.14	-0.03	0.17	0.04	0.09	0.18
Census Tracts	0.40	-0.01	-0.12	-0.08	0.04	0.11	0.03	0.06	0.20

NOTES: This table reports predicted exposure to high-income co-patrons from visiting each venue uniformly in column (1). Columns (2) through (9) shows how the realized high-income co-patron exposure of each demographic group—based on visits observed in the sample—deviates from the predicted exposure from uniform visits. The first row considers only home-venue-home visits to venues in restaurant chains. The second row considers all types of visits to venues in restaurant chains. The third row considers all types of visits to venues in all chains. The fourth row considers all types of visits to all non-residential polygons in Place IQ. The fifth row is computed as if each census tract is a venue, and individuals only visit the census tract that they live in.

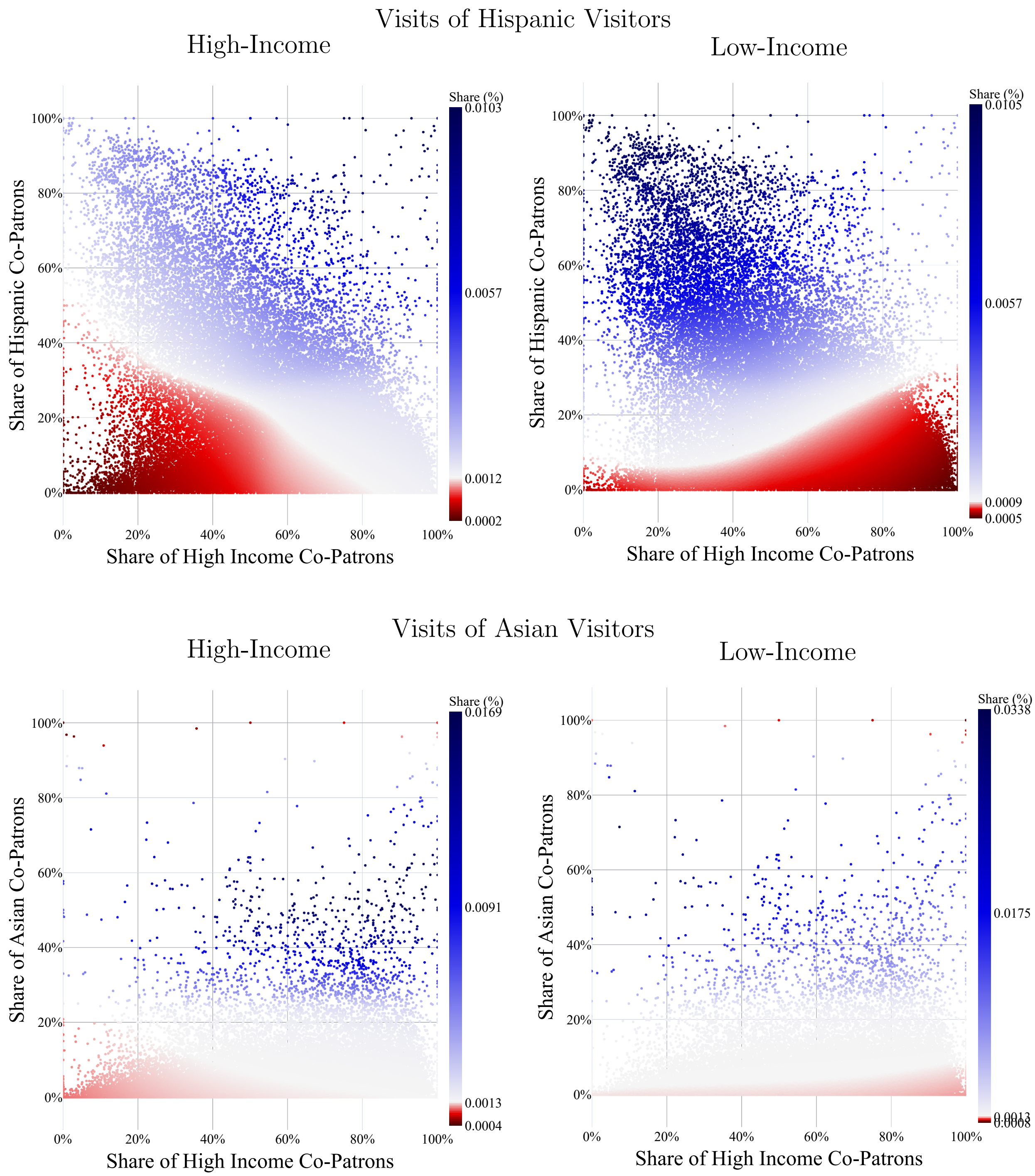
How do these differences in visit propensity and venue availability translate into differences in income exposure? Table 1 reports exposure to high-income co-patrons for each demographic group, across different environments. Each column shows the exposure of a different demographic group, relative to a baseline where people visit venues uniformly holding the characteristics of co-patrons fixed. For instance, a value of 0.10 for high-income white individuals means that the average share of high income co-patrons in the venues that

Figure 2: Exposure to (and Availability of) Co-Patron Mix



Notes: Continued on next page. This figure shows the results of a kernel regression of visit shares on co-patron race and income characteristics. Each plot shows the smoothed visit shares for a specific race by income group. Each dot corresponds to an individual restaurant venue. Shares as reported in the legend are in absolute levels. Shading of each dot is relative to the average venue, defined by the centroid venue over own-race and high-income co-patrons. The epanechnikov kernel is used with a bandwidth of 0.05.

Figure 2: Exposure to (and Availability of) Co-Patron Mix (continued)



high-income white people actually visit is 10 percentage points higher than it would be if they visited venues with uniform probability. We report our results in relative terms because absolute exposure levels are more sensitive to definitions of income group and sample biases.[10]

The first row reports income exposure for our estimation sample shown in Figure 2. We chose this sample to suit our empirical strategy, not for its representativeness, so in subsequent rows we expand our sample to more representative sets of visits and venues. Row 2 reports income exposure from all visits to restaurant chains and not just direct trips from home. Row 3 reports income exposure for all visits to commercial chain venues.[11] Row 4 reports income exposure within all non-residential establishments in PlaceIQ. Finally, row 5 reports a measure of residential income exposure, computed using only Census tables, as if people's exposure equaled the income composition of their census tract of residence. This last row offers a useful benchmark to evaluate how cross-group differences in exposure experienced within non-residential venues compare with traditional measures of residential segregation from the Census. Appendix C.1 shows a similar table for own-race instead of income exposure.

Table 1 yields two main results. First, there are substantial differences in exposure to high-income co-patrons across incomes and races. Second, we observe similar patterns and magnitudes of cross-group exposure across the different types of visits and venues that we consider. Within each race, high-income individuals have greater high-income exposure than low-income individuals: with differences in mean exposure between high- and low-income individuals generally between 15 and 20 percentage points. Within each income group, Asian and White individuals have greater high-income exposure than Black and Hispanic individuals. In fact, the average exposure of a low-income white individual is only 2 percentage points lower than that of a high-income Black individual. We find that cross-group differences in exposure to high-income peers tends to be similar in sign and magnitude in consumption venues to the neighborhoods they live in. This holds true both for differences within and across racial groups.

A number of factors may explain this variation in exposure to high-income co-patrons

[10] The patterns in Table 1 are robust to different ways of weighting each device that correct for biases in the smartphone sample. There may also be differences in exposure at the intensive margins, from variation in the number of trips that each demographic group makes. Smartphone samples with a partial history of each device's movements are not well-suited to study these differences. Travel surveys like the National Household Transportation Survey, however, show that although rich and poor individuals visit different types of destinations—for instance, rich people are more likely to visit restaurants — the overall difference in the number of trips across racial and income groups is relatively small.

[11] These include banks, big box stores, convenience store, grocery stores, gyms, pharmacies, restaurants. Results for each category shown separately in Appendix C.1.

across demographic groups. First, different groups may have differential taste for product attributes like cuisine and ambiance. Second, groups are distributed differently across cities and neighborhoods and thus might have to incur higher travel costs to patronize venues with larger high-income shares. Third, different groups may have different preferences for social exposure to the high-income co-patrons themselves. Finally, preferences for high-income and own-race co-patrons may interact with the venue environment if, as suggested in Figure 2, some racial groups lack access to own-race venues that also have high shares of high-income co-patrons. In what follows we investigate the relative importance of these explanations for differences in high-income exposure.

4 Model

This section introduces a model of individuals' decisions to patronize venues within business chains as a function of transit costs and co-patron composition. The model delivers an estimating equation for each demographic group's preferences for co-patron race and income, and costs of travel distance. We also describe how to compute counterfactual visit shares to each venue from the estimated model. These counterfactuals allow us to quantify the contributions of various mechanisms in explaining differences across demographic groups in social exposure to high-income co-patrons.

4.1 Nested-logit preferences

We develop a nested-logit model of consumers' decisions to visit venues. We index decision makers by i, venues by j, and chains by c. A decision maker is an individual at a point in time. Denote the set of venues from which a decision maker chooses by $\mathcal{J}$. The utility that decision maker i would obtain from choosing venue j is

$$U_{ij} = V_{ij} + \epsilon_{ij},$$

where V_{ij} is a scalar that depends on preference parameters and observed covariates and ϵ_{ij} is a random component. Decision maker i chooses the venue $j \in \mathcal{J}$ that has the highest value of U_{ij}.

We assume that ϵ has an extreme-value distribution such that consumers have nested-logit preferences over business chains. Following Train (2009, Ch 4.2), let the set of venues be partitioned into C disjoint subsets denoted by B_c (chains). Denote the similarity of idiosyncratic preferences for establishments in nest B_c by $1 - \lambda_c$, so that $\lambda_c = 1$ $\forall c$ is the canonical logit case. The probability that decision maker i chooses alternative j is

$$P_{j|i} = \frac{\exp\left(V_{ij}/\lambda_c\right)\left(\sum_{j' \in B_c} \exp\left(V_{ij'}/\lambda_c\right)\right)^{\lambda_c - 1}}{\sum_{c'=1}^{C}\left(\sum_{j' \in B_{c'}} \exp\left(V_{ij'}/\lambda_{c'}\right)\right)^{\lambda_{c'}}}. \tag{1}$$

4.2 Within-chain choice probabilities

If the utility shifter V_{ij} depends on preference parameters Γ, the log likelihood function associated with the choice probability (1) is

$$LL(\Gamma) = \sum_i \sum_j I_{ij} \ln P_{j|i},$$

where $I_{ij} = 1$ if i chooses j.

Following Train (2009, p.82), the choice probability (1) can be rewritten as the product of within-chain and between-chain components: $P_{j|i} = P_{j|ic} \times P_{c|i}$. Thus, we can rewrite the log likelihood function as

$$LL(\Gamma) = \sum_i \sum_j I_{ij}\left(\ln P_{j|ic} + \ln P_{c|i}\right) = \sum_i \sum_j I_{ij} \ln P_{j|ic} + \sum_i \sum_j I_{ij} \ln P_{c|i}.$$

While a model of $P_{c|i}$ must incorporate the parameters appearing in $P_{j|ic}$ via an "inclusive value" term, we can maximize the first likelihood component, $\sum_i \sum_j I_{ij} \ln P_{j|ic}$, without any constraints on the relevant elements of Γ. We estimate preference parameters of interest using only within-chain variation, leveraging the conditional choice probability

$$P_{j|ic} = \frac{\exp\left(Y_{ij}/\lambda_c\right)}{\sum_{j' \in B_c} \exp\left(Y_{ij'}/\lambda_c\right)},$$

where Y_{ij} denotes the component of V_{ij} that varies across venues within chain c.[12] We allow Y_{ij} to depend on parameters that are common across chains. In order to identify parameters common across nests, we assume a common within-chain correlation of idiosyncratic preference shocks such that $\lambda_c = \lambda\ \forall c$. In what follows, we assume that preferences for co-patron characteristics and the disutility associated with a greater distance between the consumer's

[12] Attributes that are common across establishments within a chain, such as menu items and prices, affect the inclusive value term in $P_{j|ic}$ but do not appear in $P_{j|ic}$. Any chain-level component Y_{ic} cancels out:

$$P_{j|ic} = \frac{\exp\left(\left[Y_{ij} + Y_{ic}\right]/\lambda_c\right)}{\sum_{j' \in B_c} \exp\left(\left[Y_{ij'} + Y_{ic}\right]/\lambda_c\right)} = \frac{\exp\left(Y_{ic}/\lambda_c\right)\exp\left(\left[Y_{ij}/\lambda_c\right)\right.}{\exp\left(Y_{ic}/\lambda_c\right)\sum_{j' \in B_c} \exp\left(Y_{ij'}/\lambda_c\right)} = \frac{\exp\left(Y_{ij}/\lambda_c\right)}{\sum_{j' \in B_c} \exp\left(Y_{ij'}/\lambda_c\right)}.$$

home and the venue are common across chains.[13]

4.3 Mean utility specification

In the baseline specification, we assume that the mean utility of patronizing a venue within a chain depends on transit costs and co-patron composition. These preferences may vary across demographic groups, indexed by g. Preferences over transit costs and co-patron composition are additively separable. In particular, the component of utility that varies across venues within a chain, Y_{ij}, depends on the distance from the consumer's home to the venue (distance_{ij}), the share of high-income co-patrons (s_j^{highinc}), and the share of own-race co-patrons (s_j^{ownrace}):

$$Y_{ij} = \delta^g f_1(\ln \text{distance}_{ij}) + \beta^g f_2(s_j^{\text{ownrace}}, s_j^{\text{highinc}}).$$

where δ^g and β^g are group-specific coefficient vectors on transit costs and co-patron composition, respectively, $f_1(\ln \text{distance}_{ij})$ is a polynomial of log distance with unit coefficients and $f_2(s_j^{\text{ownrace}}, s_j^{\text{highinc}})$ is a polynomial of the two co-patron shares with unit coefficients.

Choosing the degrees of the polynomials $f_1()$ and $f_2()$ involves a trade-off between parametric flexibility and statistical power. Our baseline specification uses second-degree polynomials:

$$\begin{aligned} Y_{ij} = {} & \delta_1^g \ln \text{distance}_{ij} + \delta_2^g \left(\ln \text{distance}_{ij}\right)^2 \\ & + \beta_1^g s_j^{\text{highinc}} + \beta_2^g \left(s_j^{\text{highinc}}\right)^2 + \beta_3^g s_j^{\text{ownrace}} + \beta_4^g \left(s_j^{\text{ownrace}}\right)^2 + \beta_5^g s_j^{\text{ownrace}} \times s_j^{\text{highinc}}. \end{aligned} \tag{2}$$

Second-degree polynomials are flexible enough to fit observed choice patterns well and parsimonious enough to be precisely estimated for the demographic groups with modest numbers of observations.

We can express preferences over co-patron composition in terms of willingness to travel. In particular, we can express the utility value of a co-patron composition relative to the average venue's co-patron composition as the utility value of distance to a venue relative to the distance to the average venue in the choice set. Let $(\overline{s}^{\text{ownrace}}, \overline{s}^{\text{highinc}})$ and $\overline{\text{distance}}$ denote the characteristics of the average venue. We define group g's willingness to travel for the co-patron composition $(s^{\text{ownrace}}, s^{\text{highinc}})$ as the incremental distance Δ^g that equates the mean utility of a venue at the average distance with that co-patron composition and a venue

[13] We estimate chain-specific parameters for each of the largest chains in robustness checks.

at $\overline{\text{distance}} + \Delta^g$ with the average co-patron composition:

$$\begin{aligned}
&\delta^g f_1\left(\ln \overline{\text{distance}}\right) + \beta^g f_2\left(s^{\text{ownrace}}, s^{\text{highinc}}\right) \\
&\quad = \delta^g f_1\left(\ln\left(\overline{\text{distance}} + \Delta^g(s^{\text{ownrace}}, s^{\text{highinc}})\right)\right) + \beta^g f_2\left(\bar{s}^{\text{ownrace}}, \bar{s}^{\text{highinc}}\right)
\end{aligned} \tag{3}$$

The function $\Delta^g(s^{\text{ownrace}}, s^{\text{highinc}})$ is group g's willingness to travel for that co-patron composition.[14]

4.4 Maximum likelihood estimation

We estimate the preference coefficients by maximizing the likelihood component $\sum_i \sum_j I_{ij} \ln P_{j|ic}$. The optimization problem is

$$\max_{\delta^g,\beta^g} \sum_i \sum_j I_{ij} \ln \left(\frac{\exp\left(\left[\delta^g f_1(\ln \text{distance}_{ij}) + \beta^g f_2(s_j^{\text{ownrace}}, s_j^{\text{highinc}})\right]/\lambda\right)}{\sum_{j' \in B_c} \exp\left(\left[\delta^g f_1(\ln \text{distance}_{ij'}) + \beta^g f_2(s_{j'}^{\text{ownrace}}, s_{j'}^{\text{highinc}})\right]/\lambda\right)} \right). \tag{4}$$

Since each parameter is g-specific, the model can be estimated separately by demographic group. This computation is expensive when there are many venues in the choice set and many observations, as in our smartphone data. In some cases, we exploit the independence-of-irrelevant-alternatives assumption to reduce the number of venues and observations and make the optimization problem computationally feasible.[15]

4.5 Empirical implementation

We estimate consumer preferences using 19 months of data on devices in the 100 most populous US metropolitan areas, as described in Section 2.2. We estimate the model separately by demographic group and business category, so our baseline estimates of δ^g and β^g are specific to both demographic group g and the restaurants category. We assume consumers consider all venues within their metropolitan area, so the nest B_c is the set of venues that belong to both the same business chain and metropolitan area.

We estimate consumer preferences using within-chain comparisons in order to distinguish consumer preferences over co-patron composition from other venue characteristics.

[14] The equation implicitly defines this function. Given the functional form used in equation (2), there is a closed-form expression for $\Delta^g(s^{\text{ownrace}}, s^{\text{highinc}})$.

[15] This applies in a small number of cases: we only need to reduce the dimensionality of the problem when estimating the preferences of the largest demographic groups in the largest cities patronizing chains with many venues. The reduction procedure, introduced by McFadden (1978), is described further in Davis et al. (2019). We apply it when a chain has more than 75 venues in a city or when we observe more than 20,000 decision events by a demographic group in a city.

Co-patron composition may correlate with other traits in the very broad set of (potentially unobserved) characteristics entering V_{ij}, such as service quality, comfort, or product offering. The set of characteristics contributing to Y_{ij}, which vary across venues within a chain business and metropolitan area, is substantially smaller. Venues in the same chain typically offer similar service quality, comfort, and products. In robustness checks, we focus on a subset of the most uniform chains based on franchising terms and dispersion in establishment-level characteristics.

Our estimation of consumer preferences using the likelihood function (4) predicts patronage decisions as a function of bilateral distance and co-patron composition. There is no mechanical sense in which co-patron composition will predict group-specific patronage decisions because bilateral distance is an individual-by-venue-specific cost shifter.[16] Moreover, our estimation sample of direct trips from home is a small share of the total visits to venues, so the observed co-patron composition is not driven by the choice shares in our estimation sample.

For brevity, we refer to all mechanisms that cause co-patron composition to predict consumer decisions as preferences over co-patron composition. Of course, co-patron composition may predict decisions not because consumers have preferences over co-patron demographics but because co-patron demographics predict other elements of the decision. For example, a consumer who is indifferent to strangers' demographics may choose a venue in order to meet up with their demographically similar friends.[17] This behavior could generate the same observed outcomes as a consumer who has homophilic preferences over anonymous co-patrons. We need not separate homophily among strangers and homophily in friendship networks to quantify the importance of homophily in explaining cross-group differences in experienced income exposure. Similarly, if consumers are not aware of all the venues in their choice set, co-patron demographics could predict consideration sets. Our estimation approach would infer homophilic preferences if demographically similar venues are more likely to be considered. While this distinction would be important when considering some counterfactual scenarios, our decomposition of observed social exposure to high-income co-patrons will not distinguish preferences over co-patron demographics from consideration sets that vary with demographics.

[16]We could extend our venue choice model to account for residential choice as in the Davis et al. (2019) appendix. This extension shows that residential sorting does not bias estimates of preferences for venues as long as individuals choose a residence based on the expected utility of their venue choice set (and not on the utility of any specific venue).

[17]We also replicate our preference estimation in categories other than restaurants — like banks and big box stores — where meeting friends is less likely.

4.6 Decomposition of social exposure to co-patron characteristics

We decompose social exposure to high-income co-patrons by contrasting the distribution of visits across venues in our fitted model with the distributions resulting from various counterfactual market shares. We summarize social exposure to high-income co-patrons for members of group g by fitting a density $f^g(\cdot)$ using kernel $K(\cdot)$ and bandwidth h to the high-income share in each venue:

$$\hat{f}^g\left(s^{\text{highinc}}\right) = \frac{1}{h}\sum_{j\in\mathcal{J}} K\left(\frac{s^{\text{highinc}} - s_j^{\text{highinc}}}{h}\right) P_{j|g}. \tag{5}$$

To compute our benchmark 'model-predicted' distribution of social exposure to high-income co-patrons, we define the share of visits to each venue j by group g as follow:

$$P_{j|g} = P_{j|ic} \times P_{ic|g}, \tag{6}$$

where $P_{j|ic}$ comes from our estimated model of within-chain venue choice, and $P_{ic|g}$ comes from the observed distribution of visits to each chain by residential origin among members of a demographic group.

To quantify the contributions of various mechanisms to social exposure, we report the distributions resulting from various counterfactual market shares $P_{j|g}$. A simple starting point is the observed distribution of high-income co-patrons across all venues. This is the density that results from evaluating equation (5) using a uniform probability of visiting venues, $P_{j|g} = \frac{1}{|\mathcal{J}|}$. To quantify the contribution of between-CBSA variation in demographics to social exposure to high-income co-patrons, we use a uniform probability conditional on metropolitan area, $P_{j|g} = \frac{1}{|\mathcal{J}_m|}P_{m|g}$. The difference between the nationwide uniform probability $\frac{1}{|\mathcal{J}|}$ and the measure that reflects which metropolitan areas group members tend to reside captures the contribution of CBSA sorting to social exposure. To quantify the contribution of between-chain variation in demographics, we use a uniform probability conditional on metropolitan area and business chain, $P_{j|g} = \frac{1}{|\mathcal{J}_{mc}|}P_{mc|g}$. To quantify the contribution of neighborhood sorting, we compute market shares with counterfactual probabilities $P_{j|ic}$ using the estimated distance parameters $\hat{\delta}^g$ absent any co-patron preferences ($\beta^g = \mathbf{0}$). To quantify the contribution of preferences over co-patron composition, we compute market shares with counterfactual probabilities $P_{j|ic}$ using the estimated co-patron preference parameters $\hat{\beta}^g$ absent any disutility of distance ($\delta^g = \mathbf{0}$).[18]

[18] The last two counterfactuals in which we set $\beta^g = \mathbf{0}$ or $\delta^g = \mathbf{0}$ are non-nested scenarios. Nested scenarios, such as the stratified uniform-probability scenarios, lend themselves to simple comparisons because they differ in only one respect. Non-nested scenarios, such as alternatively setting preference parameters for distance or

5 Estimation results

In this section, we estimate willingness to travel for social exposure to co-patrons within restaurant venues. We document notable regularities in social preferences: different income and racial groups display similar levels of racial homophily. Preferences for high-income co-patrons are also similar across racial groups, but lower income individuals have less pronounced tastes for co-patron income. These preference patterns are consistent across a number of robustness checks, including restricting our estimation sample to the most standardized restaurant chains.

5.1 Estimated preference parameters

Table 2 reports estimates of the preference parameters in equation (2) for each of the eight demographic groups. Panel A reports estimates of the distance coefficients δ^g, and Panel B reports estimates of the co-patron composition coefficients β^g.

The estimates of δ^g imply distance elasticities around -2.2. Evaluated at group-specific average visit distances, ranging from 5.7 to 7.2 km, the distance elasticities span -2.10 to -2.36 across the eight demographic groups. These distance elasticities capture both the cost of longer travel distances, and the substitutability of venues within a restaurant chain. Higher-income individuals have higher distance elasticities, consistent with empirical evidence that the value of time spent traveling rises with income (Small and Verhoef, 2007). Our distance elasticities are larger than previous estimates from studies that consider venue choice amongst all restaurant, consistent with venues within the same chain being closer substitutes.[19]

Table 2 Panel B reports, for each of the eight demographic groups, estimates of all five coefficients in β^g, which together govern preferences for the share of high-income co-patrons and the share of own-race co-patrons. In Figure 3 and 4, we propose two different visual representations of these preferences. Figure 3 depicts social preferences in detail, over both the income and race of co-patrons, with the preferences of each demographic group in a separate heatmap. Each point represents a chain restaurant venue in the 100 largest metropolitan areas. The horizontal axis shows the share of high-income co-patrons within a venue, and the vertical axis shows the share of own-race co-patrons. The color of the venue captures willingness to travel to that venue in co-patron composition space,

for co-patron composition to zero, must be interpreted carefully. These alternative scenarios do not provide an additive decomposition of the observed outcomes, because marginal effects are non-linear functions of the coefficients and covariates. We address this issue further in our discussion of these results in Section 6.

[19] Athey et al. (2018); Davis et al. (2019); Couture et al. (2023) find elasticities between -1.0 and -1.5 when considering substitution between all restaurants or non-tradable services in a given city.

Table 2: Preference estimates

Panel A. Estimates of distance coefficients δ^g

		Estimates		Distance Elasticity at				
		Linear δ_1^g	Quadratic δ_2^g	Mean	10p	25p	75p	90p
Low-Income	White	-1.11	-0.32	-2.30	-1.01	-1.46	-2.44	-2.86
Low-Income	Black	-1.25	-0.23	-2.10	-1.19	-1.53	-2.22	-2.49
Low-Income	Hispanic	-1.26	-0.27	-2.19	-1.12	-1.49	-2.29	-2.64
Low-Income	Asian	-1.29	-0.23	-2.10	-1.14	-1.48	-2.18	-2.50
High-Income	White	-1.00	-0.36	-2.36	-1.12	-1.56	-2.51	-2.95
High-Income	Black	-1.11	-0.30	-2.31	-1.22	-1.61	-2.45	-2.82
High-Income	Hispanic	-1.22	-0.30	-2.31	-1.23	-1.60	-2.41	-2.81
High-Income	Asian	-1.25	-0.26	-2.21	-1.27	-1.60	-2.30	-2.65

Panel B. Estimates of co-patron composition coefficients β^g

		β_1^g Linear High-Income	β_2^g Quadratic High-Income	β_3^g Linear Own race	β_4^g Quadratic Own race	β_5^g Interaction Term
Low-Income	White	2.833	-4.134	2.262	-1.793	2.46
		(.072)	(.059)	(.086)	(.075)	(.087)
Low-Income	Black	5.129	-4.236	4.977	-3.867	-2.287
		(.065)	(.054)	(.053)	(.045)	(.064)
Low-Income	Hispanic	6.048	-4.199	7.077	-4.977	-4.217
		(.07)	(.054)	(.076)	(.065)	(.068)
Low-Income	Asian	4.125	-3.195	12.194	-9.929	-4.016
		(.078)	(.07)	(.119)	(.155)	(.135)
High-Income	White	4.159	-3.595	4.434	-3.631	2.556
		(.086)	(.066)	(.113)	(.09)	(.096)
High-Income	Black	5.402	-3.215	4.17	-3.61	-.327
		(.071)	(.053)	(.052)	(.044)	(.053)
High-Income	Hispanic	7.827	-4.24	7.199	-5.114	-4.524
		(.087)	(.06)	(.091)	(.078)	(.076)
High-Income	Asian	5.685	-3.314	12.798	-12.281	-2.274
		(.067)	(.051)	(.085)	(.099)	(.092)

NOTES: This table reports estimates of the preference parameters in equation (2). Panel A reports estimates of the distance coefficients δ^g, and Panel B reports estimates of the co-patron composition coefficients β^g. The average trip distance for each demographic group in the estimation sample ranges from 5.7 to 7.2 kilometers. The 10th and 90th percentiles of trip distance range from 0.7 to 1.2 and from 13.6 to 16.4 kilometers, respectively.

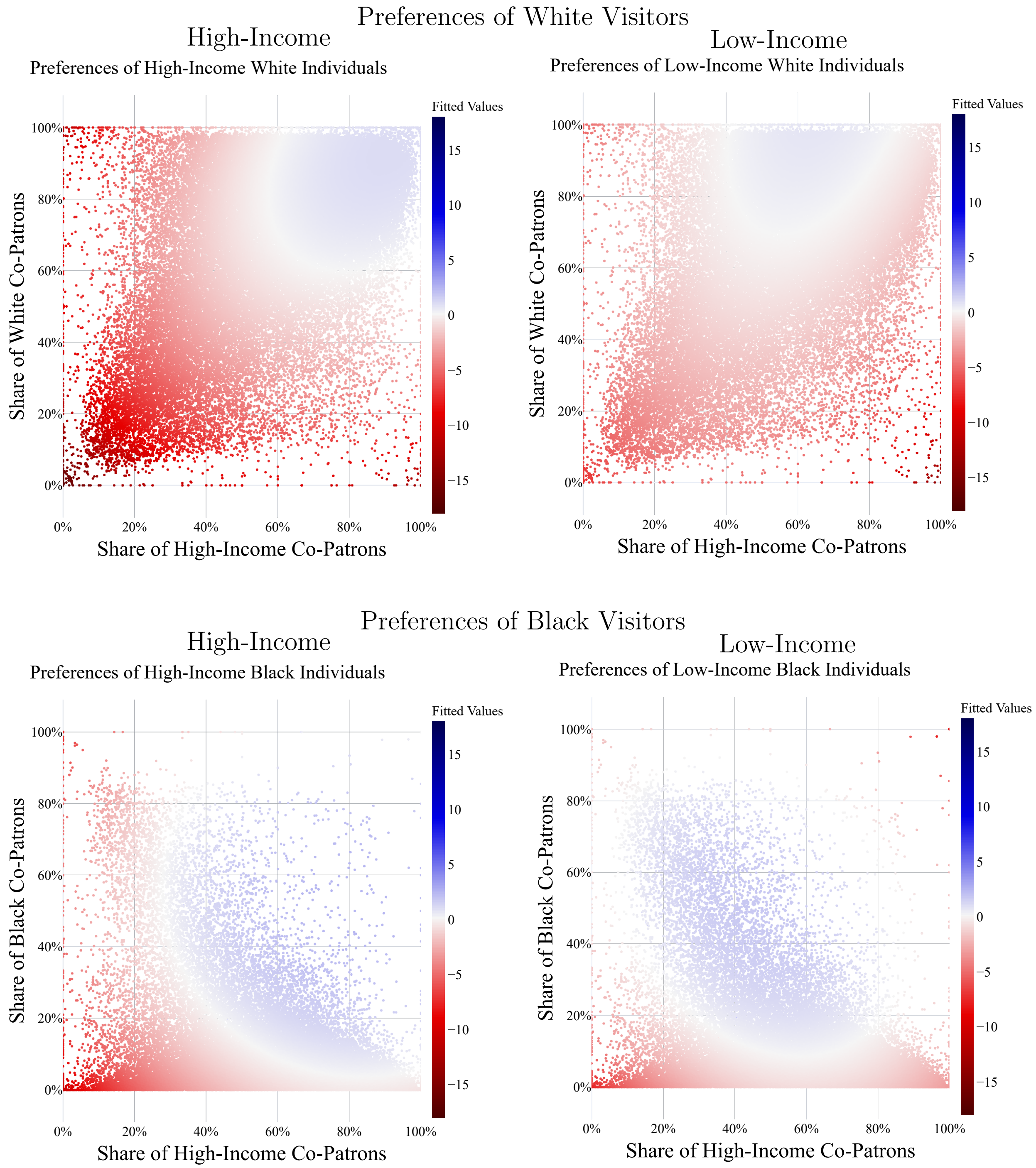
$\Delta^g(s^{\text{ownrace}}, s^{\text{highinc}})$ from equation (3). Venues with co-patron demographics preferable to the average co-patron composition are depicted in blue; those worse than average are in red. To facilitate comparisons across groups, Figure 4 depicts the preferences of all eight demographic groups over one dimension of co-patron composition on the same plot, fixing the other dimension at its median value for each group. This is akin to looking at variation along one horizontal or vertical slice of Figure 3.

We find that preferences for co-patron income are remarkably similar across racial groups, but high-income individuals have a stronger taste for high-income co-patrons. Figure 3 shows that high-income individuals have monotone preferences over high-income co-patron share, increasing from left to right in each plot, whereas their low-income counterparts have non-monotone preferences for co-patron income, with the darkest shades of blue being in the middle of the plot. Figure 4 Panel A quantifies the strength of income preferences by showing variation in willingness to travel for only establishments with the median own-race share for each group. Across all racial groups, high-income individuals are willing to travel 2.4 to 3.4 additional kilometers to visit a venue in the top decile, rather than the bottom decile. Lower income individuals have less pronounced income preferences, with the most preferred share of high-income co-patron between 50 to 60 percent for all racial groups. Low-income Black, Hispanic and Asian individuals are willing to travel around 1.2 additional kilometers to visit a venue with that optimal income composition (relative to a venue in the bottom decile), while low-income White individuals have an even lower willingness to travel around 0.6 kilometer.

Turning to preferences for own-race co-patrons, we find that all demographic groups exhibit substantial racial homophily. In all eight panels of Figure 3, the strongest preferences are for establishments in the venues offering relatively high own-race shares. Figure 4 Panel B shows that the strength of this racial homophily does not vary by income. Comparing levels of racial homophily across racial groups is harder because preferences are estimated over different support in terms of own-race share. White individuals are the racial majority in most venues, while other races have few venues in their choice sets with large own-race shares. That said, White, Black and Hispanic individuals have similar own-race preferences across deciles; all are willing to travel about 2 km farther to visit a venue in the top rather than bottom decile of own-race share. Asian racial homophily appears stronger than that of the three other groups, albeit on very limited support. Finally, we note that racial and income preferences are roughly similar in magnitude. This similarity turns out to be important for how people trade-off income exposure for racial exposure, as we show in the next section.

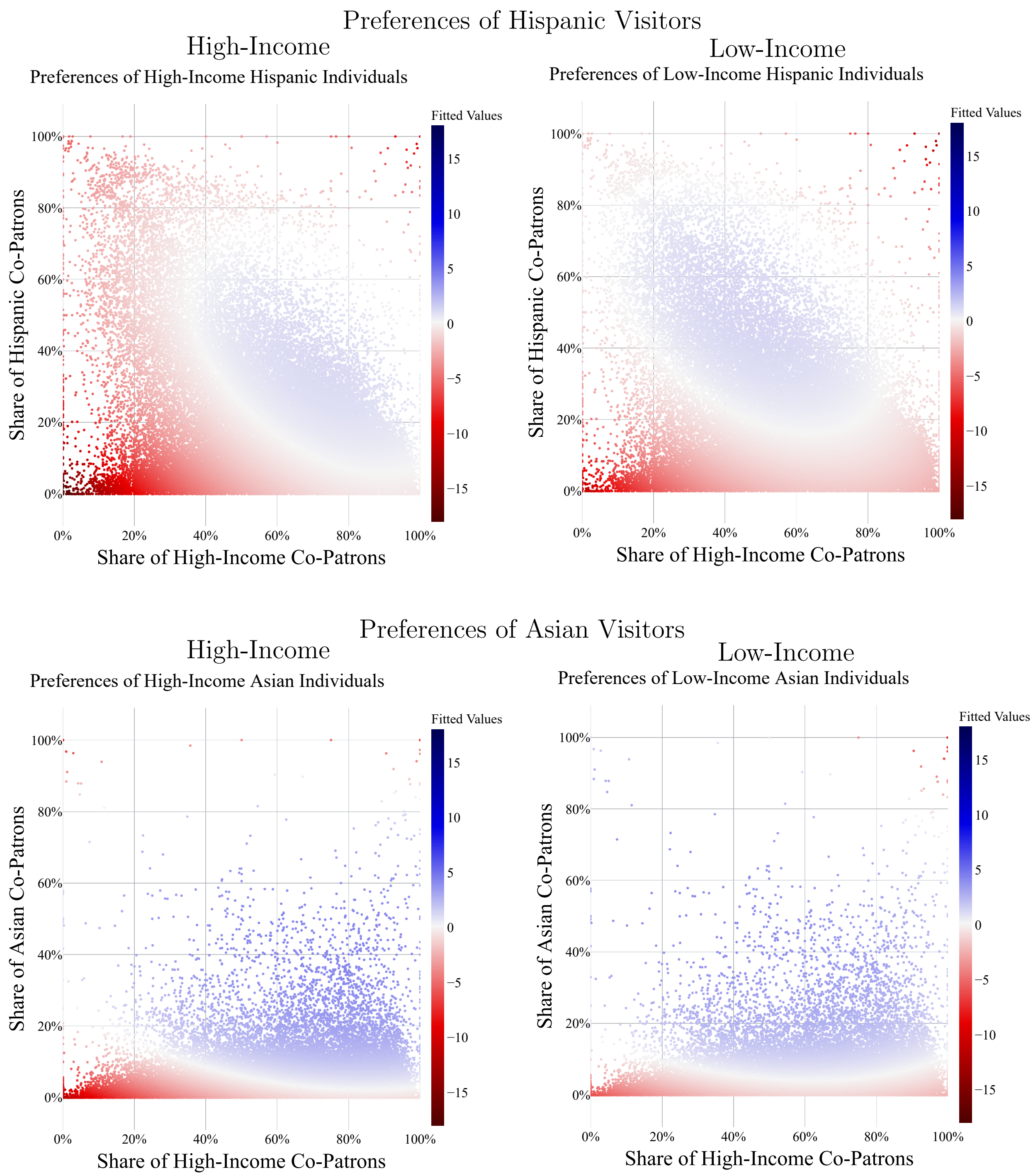
Our presentation in this paper emphasizes broad patterns of social preferences, but our detailed preference representation could offer much additional insight into how people choose

Figure 3: Preferences over co-patron demographics



Notes: Continues onto next page. Each plot shows the co-patron valuation of chain restaurants given by equation (2) for each of the eight race by income groups. Each dot corresponds to an individual venue. Co-patron valuations are adjusted relative to the average venue, defined by the centroid venue over own-race and high-income co-patron shares.

Figure 3: Preferences over co-patron demographics (continued)



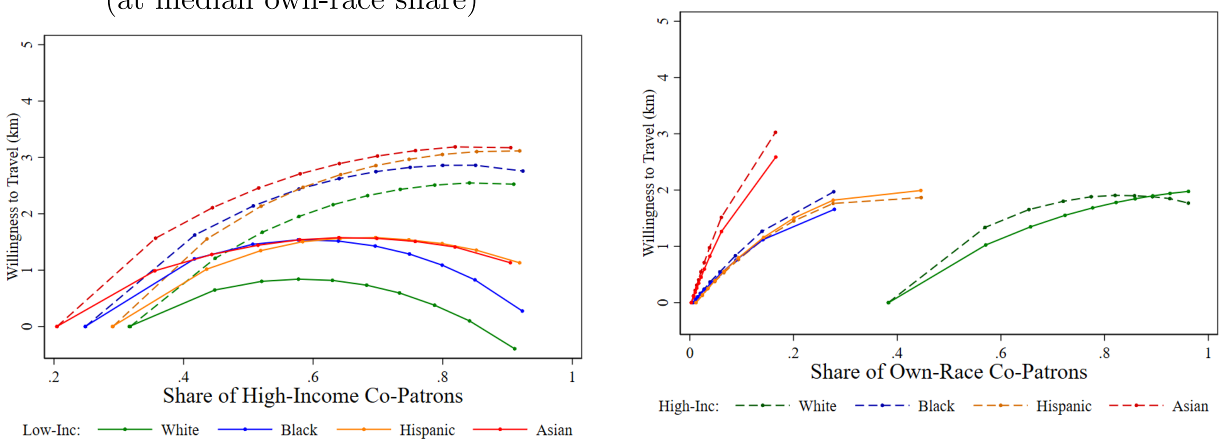
their social exposure. Here, we briefly discuss two additional features of these preferences. First, Figure 4 shows that social preferences are non-linear. That is, when choosing between venues with few own-race or few high-income co-patrons, people display a strong willingness to travel to higher income and own-race shares. Starting from already high shares of own race or high-income co-patrons however, we observe much less willingness to travel to venues with even higher shares. This suggests that social preferences might partly stem from an aversion to being in an overwhelming racial minority, and to visiting venues heavily patronized by poor people. Second, there are complementarities between race and income in preferences for social exposure. Such complementarities are visible in Figure 3, but easier to assess by looking directly at the interaction coefficients in Table 2. We estimate positive interaction terms for high- and low-income White individuals and negative interaction terms for all other groups. White individuals care more about the share of high-income patrons at higher shares of own-race co-patrons. For all other racial groups, the reverse is true: co-patron income matters less as own-race share rises. From the perspective of non-White individuals, the share of co-patrons who are not own-race is composed mostly of White co-patrons, so these preferences suggest that all four racial groups have sharper income preferences when co-patrons are White.[20]

[20]Seven out of eight interaction terms are between 2.3 and 4.5 in absolute value. To offer a sense of magnitude, the coefficient of -2.3 for low-income Black individuals implies that when visiting an establishment in the top quartile of the Black co-patron share, they are willing to travel an extra 0.65 kilometers to visit a higher-income venue (at the 90th percentile of high-income share relative to the 10th percentile of high-income share.) When visiting an establishment with fewer Black co-patrons, in the bottom quartile of Black co-patron share, low-income Black individuals are willing to travel even longer to visit a higher-income venue, an additional 1.84 kilometer.

Figure 4: Preferences for High-Income and Own-race Co-Patrons

A. Willingness to Travel for High-Income Share (at median own-race share)

B. Willingness to Travel for Own-Race Share (at median high-income share)



NOTES: This figure depicts the preference estimates reported in Table 2 Panel B along each dimension of co-patron composition while fixing the other. Panel A shows preferences over the share of high-income co-patrons when the share of own-race co-patrons is fixed at the median. Panel B shows preferences over the own-race share while fixing income composition. The ten points in each series are plotted at the mean of each decile of the co-patron shares observed in the estimation sample. The median White share is 77%; the median Black share is 3%; the median Hispanic share is 9%; the median Asian share is 2%.

Translated to dollars, these parameter estimates imply substantial willingness to pay for preferred social exposure. At typical travel speeds and values of time, an additional kilometer translates to about one dollar, so high-income individuals traveling 2.5–3.0 km farther to visit a venue in the top decile of high-income co-patron share rather than the bottom decile implies a \$2.50–\$3.00 difference in willingness to pay per trip.[21] The 2km difference between the top and bottom deciles of own-race share for Black, White and Hispanic individuals of both income groups translates to a \$2 difference per trip. These estimates imply a \$1,000-1,500 annual willingness to pay to span the range of available social exposure along these margins, as the average US driver makes more than 500 consumption trips per year (Couture, Duranton, and Turner, 2018). For context, Black (1999) estimates that the marginal resident is willing to pay approximately 2.1 percent of the mean house price to access schools with one standard deviation higher test scores, which amounts to \$3948, amortized over the years during which one lives in that house. Like school quality, social preferences may therefore be an important determinant of neighborhood choice.

[21] These calculations assume households make roundtrips from home at an average speed of about 40km/hour (Couture, Duranton, and Turner, 2018) and that they value time at \$19 per hour (Goldszmidt et al., 2020).

5.2 Robustness

This section addresses two potential sources of estimation bias: within-chain heterogeneity in venue characteristics and mis-specification of the travel-cost function $f_1(\ln \text{distance}_{ij})$. Restaurant chains offer standardized settings and products, but there still may be within-chain variation in venue characteristics that correlate with co-patron composition. To address this, we estimate specifications in which we (i) restrict the estimation sample to the chains with the most standardized venues and (ii) control for more venue and neighborhood characteristics. Travel costs might also differ from a quadratic polynomial in log distance in a way that is correlated with co-patron composition. To address this, we estimate specifications in which we (i) restrict the estimation sample to cities in which trips are overwhelmingly made by car and (ii) use more flexible functions of distance.

To facilitate comparisons across samples and specifications, we use a parsimonious specification in which $f_2(s_j^{\text{ownrace}}, s_j^{\text{highinc}})$ is a first-order polynomial so that preferences over co-patrons are a single coefficient for co-patron income and a single coefficient for co-patron race. Broadly, these robustness checks deliver preference estimates that are quite similar across the various samples and specifications.

5.2.1 Within-chain heterogeneity

We restrict attention to the most standardized chains based on two chain characteristics. The first is the coefficient of variation (CV) in the Google Places star rating across venues within the chain.[22] Less variation in reviewer ratings across venues suggests a more standardized service. The second characteristic is ownership structure. Following Williamson (1991)'s argument that franchising facilitates local adaptation, we expect chains with franchisees to be less standardized than owner-operated chains.[23] Out of 76 restaurant chains in our sample, we classify the 10 chains with fewer than 5 percent franchised venues as "entirely wholly-owned" chains, the 7 chains with between 5 and 20 percent of franchised venues as "almost wholly-owned", and the remaining chains as "franchised."[24]

These two measures of chain standardization are consistent with one another. The av-

[22]The Google Place data on restaurant venue location and characteristics comes from Akbar et al. (2023). We were able to match 41 percent of PlaceIQ venues in our estimation sample to a venue in that Google Place data. Appendix A provides more details on the Google Place data and variable construction.

[23]Krueger (1991) makes the related argument that franchisees may shirk on quality by free-riding on brand reputation. In a meta-analysis of 44 studies on franchising, Combs and Ketchen Jr (2003) find support for the hypothesis that agency theory explains franchising. For instance, more geographically dispersed chains have higher franchising rates.

[24]We collect the franchise data from multiple sources: annual reports to investors for the 2020 fiscal year, company websites, franchise disclosure documents, and franchise database compiled by Entrepreneur magazine.

erage Google Places star rating variation of a franchised chain is twice is large as that of a wholly-owned chain. Four of the five chains with the least variation in star ratings are wholly-owned, and all ten entirely wholly-owned chains are within the top third of chains with the least variation in star ratings. None of the largest and perhaps most familiar chains – like McDonalds, Subway, Starbucks, and Burger King – are particularly standardized using these metrics.[25]

Figure C.11 reports estimation results for all eight demographic groups within five different samples of restaurant chains: the baseline sample with all chains, only entirely wholly-owned chains, only almost wholly-owned chains, and the bottom quartiles of chains (weighted by number of venues) with the lowest coefficient of variation for star rating and square footage, respectively.[26] The preference estimates are qualitatively similar across all these chain samples, albeit noisy for some groups due to small visit counts. Overall, the preference patterns highlighted in Section 5.1 hold within the most standardized restaurant chains, which are less subject to concerns about variation in service quality and menu across venues.

Our second set of robustness checks addresses within-chain heterogeneity by controlling for more venue and neighborhood characteristics. Figure C.12 depicts the results of adding three venue characteristics: the Google Places star rating, the Google Places number of reviews, and the venue square footage from PlaceIQ. Adding these covariates has little impact on estimated preferences for own-race co-patrons. It increases willingness to travel for a larger share of high-income co-patron across all eight demographic groups. Finally, we control for the demographic composition of the residents of or visitors to the neighborhood in which the venue is located. The coefficients on the co-patron composition of the venue itself are similar in sign and magnitude when we add controls for the shares of own-race and high-income residents in the venue's census tract or the shares of own-race and high-income co-patrons within all other commercial venues located in a venue's census tract (Figure C.13).[27]

[25]McDonalds, Subway, Starbucks, and Burger King all fall within the top third of chains with the highest star rating CV, and all are franchised (except Starbucks, which is hybrid with about 45 percent of franchised venues as of September 2019 based on the company's 10-K filing). The five chains with the lowest coefficients of variation are Culvers, Longhorn Steakhouse, Olive Garden, MOD Pizza, and Cracker Barrel. Of these, only Culvers is franchised.

[26]We include variation in square footage as an additional robustness check, but these data are often mismeasured (it sometimes includes parking lots for instance) and we are not confident that it captures true variation across chains.

[27]The only exception is the income preferences of low-income individuals, which become weaker after adding controls for visitors to other venues in the same census tract. The income preferences of low-income individuals are quadratic and hardest to capture with a single coefficient, so these coefficient estimates are less stable.

5.2.2 Travel-cost specification

The preference estimates reported in Section 5.1 could be biased by any component of travel costs not predicted by a quadratic polynomial in log distance, and correlated with co-patron composition. For example, venues with more high-income co-patrons may be in locations better accessed by walking rather than driving. Or individuals may have a special taste for very short trips, which would tend to be to demographically similar venues, given residential sorting by income and race. We address these concerns by restricting attention to car-dominated cities and using more flexible functions of distance.

To address varying transport-mode choices, Figure C.14 reports preference estimates for the eight demographic groups for subsets of the 100 largest MSAs based on car usage. The preference estimates for all 100 MSAs are similar to those obtained when restricting the estimation sample to MSAs in which at least 90% or 95% of trips to commercial venues are by car.[28]

Figure C.15 reports preference estimates using two alternative functions of distance. The first uses a cubic polynomial of log distance, which is more flexible than our baseline quadratic polynomial. The second introduces a dummy variable indicating the venue closest to the individual's residence, which would capture a preference for very short trips or a particular salience of the nearest establishment. These specifications both yield coefficients on co-patron shares very similar to our baseline specification.

A final piece of evidence suggesting that the travel-cost function is well specified comes from event studies of moves between demographically distinct neighborhoods reported in Section 6 below. If preference estimates were biased by neighborhood demographics co-varying with distances to venues, estimated preferences would shift when individuals move between neighborhoods with different demographics. Figure C.16 does not show any discontinuous shift in preferences around such moves.

5.2.3 Other categories of commercial venues

Our baseline analysis reported results for the largest venue category, restaurant chains. Restaurants and coffee shops have been singled out as a plausible setting for social exposure by other studies (Athey et al., 2021; Atkin, Chen, and Popov, 2022; Massenkoff and Wilmers, 2023), and restaurant chains generally strive to provide a consistent experience across venues within the same city. It is possible, however, to estimate social preferences within other kind of commercial venues that have chains. A priori, it is unclear whether to

[28]These MSA-level statistics come from the 2017 National Household Travel Survey. Almost 90% of trips to commercial venues in the United States are by car. Appendix A provides more detail on the NHTS data and variable construction.

expect weaker or stronger preference estimates in other settings. For instance, preferences for co-patrons within retail chains may be weaker if social exposure is less salient in a store environment. Conversely, preference estimates may be biased upward if stores tailor their product offering to the characteristics of their clientele, for instance by offering higher quality products in richer neighborhoods, or more shelve space for Asian food in a predominantly Asian neighborhood. Figure C.10 reports estimated preferences for each demographic group for eight distinct venue categories: banks, big-box stores, convenience/gas stores, grocery stores, gyms, pharmacies, restaurants, and all business categories pooled together.[29] We find some variation in the magnitude of preferences across categories, but the results confirm that the visit patterns documented in Section 5.1 are not unique to restaurants. Within all business categories, people preferentially visit venues with more co-patrons of their own-race. People also preferentially visit venues with more high-income co-patron, with this inclination being again more muted for low-income people. A notable exception is banks, where low-income individuals avoid branches with high-income co-patrons. This exception is unsurprising: given the nature of banking services, different branches of the same bank must tailor their services and advisory expertise to the income of their clientele.

6 Determinants of social exposure

This section examines how preferences over co-patron demographics contribute to realized social exposure to high-income individuals. The fact that, given their residences, individuals choose establishments with different co-patron demographics has a direct effect on the resulting social exposure. Section 6.1 shows that differences in residential locations and differences in co-patron preferences are each sufficient to singlehandedly explain the exposure to high-income co-patrons experienced by high-income Asian, Hispanic, and White individuals. For high-income Black individuals, racial homophily dominates their preference for high-income co-patrons, but residential location explains their low social exposure to high-income co-patrons more than preferences over co-patron demographics. Beyond their direct effects, preferences over co-patron demographics turn out to be informative about residential sorting. Section 6.2 shows that, within demographic groups, individuals reside in and move to neighborhoods with demographics that are positively correlated with their preferences over co-patron demographics. This relationship suggests that preferences over both neighbors' and co-patrons' demographics are aligned or there are mechanisms that link preferences over co-patrons to residential experiences.

[29] Table A.5 shows the five largest chains in each category.

6.1 Decomposition of exposure to high-income co-patrons

Beyond the differences in preferences estimated in Section 5, there are differences in the characteristics of choice sets that generate differences in exposure to high-income co-patrons across demographic groups. In particular, choice sets differ in three dimensions. First, individuals live in different CBSAs, which offer different venue choice sets. Second, individuals visit different chains, which attract different mixes of co-patrons. Finally, travel is costly, and individuals live in neighborhoods in which nearby venues have different co-patron composition. Since individuals sort into CBSAs, chains, and neighborhoods by income and race, these differences could explain heterogeneity in high-income exposure across demographic groups. Preferences play a role in driving the differences in exposure conditional on access and in driving the sorting behavior that generates differences in access. To quantify each of these components, we compute model-predicted visit densities under the different counterfactual scenarios described in Section 4.6.

Table 3 presents the results of this decomposition of the mean exposure to high-income co-patrons.[30] Each cell describes the visit-weighted average share of high-income co-patrons for the demographic group given by the column in the counterfactual scenario given by the row. The "Venue" row assumes that all venues are visited uniformly; the outcome is the same for all groups and equal to the 60% unweighted average high-income co-patron share across venues. The subsequent rows show changes in the visit-weighted average share of high-income co-patrons relative to this 60% benchmark. The complete model specification ("Model-predicted visits") yields results very close to the exposure observed in the data ("Estimation-sample visits").

The intermediate rows of Table 3 introduce different types of sorting to illustrate why, facing the same national venue distribution, different groups of individuals experience substantially different income exposure. The CBSA sorting and chain sorting rows reveal how important the sorting of individuals into CBSA-chain nests is for explaining differences in exposure. Then, conditional on the CBSA-chain nest within which we model choices of venues, we compare how factors accounted for in the model—proximity and social preferences—determine model-predicted visits.

Role of CBSA-chain nest The "CBSA sorting" row depicts the change in mean exposure to high-income co-patrons from the 60% benchmark if individuals visited venues uniformly within their CBSA of residence. Accounting for CBSA sorting barely shifts exposure relative to the benchmark. The largest shift is for low-income Hispanic individuals, who tend to

[30]The decompositions of the full density for each group are available in Appendix C.3.

Table 3: Mean exposure to high-income co-patrons

	Low Income				High Income			
	White (1)	Black (2)	Hispanic (3)	Asian (4)	White (5)	Black (6)	Hispanic (7)	Asian (8)
Venue	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60
CBSA sorting	-0.01	-0.01	-0.05	-0.02	0.01	0.01	-0.02	0.01
CBSA-chain sorting	-0.00	-0.02	-0.06	-0.02	0.04	0.02	-0.01	0.04
Within-nest differences								
Preference sorting	0.01	-0.05	-0.07	-0.00	0.10	0.04	0.03	0.09
Income preference sorting only	-0.05	-0.01	-0.01	0.00	0.04	0.06	0.07	0.09
Race preference sorting only	-0.00	-0.08	-0.10	-0.00	0.04	-0.01	-0.05	0.05
Neighborhood sorting	-0.03	-0.16	-0.15	-0.06	0.11	-0.01	0.04	0.11
Model-predicted visits	-0.03	-0.16	-0.15	-0.06	0.13	-0.00	0.05	0.14
Estimation-sample visits	-0.02	-0.16	-0.15	-0.06	0.14	-0.00	0.05	0.14
Actual visits (All)	-0.02	-0.13	-0.13	-0.04	0.12	-0.00	0.04	0.12

NOTES: This table shows the average share of high-income co-patrons an individual of each group would be exposed to under different counterfactual visit scenarios. All rows except the first row are adjusted relative to the "Venue" row. The first row evaluates the counterfactual scenario when devices visited venues uniformly at the national level. The second row evaluates the counterfactual scenario when devices visited venues uniformly within their CBSA of residence. The third row evaluates the counterfactual scenario when devices visited venues uniformly within their CBSA of residence and choice of chain. The fourth row evaluates the counterfactual scenario when devices consider their preferences for co-patron characteristics, in addition to their choice of CBSA and chain. The fifth row evaluates the counterfactual scenario when devices consider their distance dis-utility while ignoring preferences for co-patron characteristics, in addition to their choice of CBSA and chain. The sixth row evaluates the counterfactual scenario when devices only consider their preferences for high-income co-patrons, in addition to their choice of CBSA and chain. The seventh row evaluates the counterfactual scenario when devices only consider their preferences for high-income co-patrons, in addition to their choice of CBSA and chain. The eighth row evaluates the counterfactual scenario when devices consider both their preferences for co-patron compositions and distance dis-utility, in addition to their choice of CBSA and chain. The ninth row shows the actual exposure based on the home-venue-home visits in our estimation sample. Finally, the tenth row shows the exposure resulting from all types of visits.

reside in poorer cities. For them, differences across CBSAs explain only one-third of their experienced income exposure relative to the benchmark (-0.05 of -0.15).

The "Chain sorting" rows reports the change in mean exposure from the benchmark if individuals visited venues uniformly within the chains that they visit in the CBSA where they reside. Accounting for chain sorting has little effect: most exposure differences stem from within-chain sorting, despite the differences in visit propensities across income and racial groups presented in Figure A.2. The largest shifts in exposure are among high-income White and Asian individuals, whose choice of chain shifts their high-income exposure up by 3 percentage points, less than one-quarter of the overall difference between their experienced exposure and the national average.[31] Differences in the distribution of groups across cities and chains fail to explain most of the differences in income exposure across groups.

Within-nest differences: Proximity versus preferences Most of these differences in income exposure must be explained by either relative proximity or social preferences. The "Preference sorting" and "Neighborhood sorting" rows of Table 3 summarize the impact of each factor on predicted income exposure. Specifically, the "Preference sorting" row depicts the relative mean exposure if venues within each chain-CBSA nest were visited only based on preferences for co-patron composition absent any disutility of distance ($\delta^g = \mathbf{0}$). Conversely, the "Neighborhood sorting" row shows the income exposure outcomes that would result if individuals picked which of a chain's establishments to visit based only on distance parameters $\hat{\delta}^g$ absent any co-patron preferences.

Accounting for neighborhood sorting beyond CBSA-chain sorting sharply decreases the mean predicted high-income exposure of low-income individuals of all groups. The neighborhoods in which low-income individuals reside are the most important factor in explaining why they have less than the average high-income co-patron exposure. Preferences, by contrast, are far less powerful at predicting this difference: the income exposure experienced by low-income individuals is much lower than what their social preferences alone would suggest. This is consistent with lower-income individuals being unable to afford high-income neighborhoods.

We see a similar divergence in the impact of preferences and neighborhood sorting on our predicted high-income exposure of high-income Black individuals. If high-income Black individuals chose which venues to visit based on their preferences, their high-income exposure would be higher than it is in the data. This tendency is counteracted by the impact of neighborhood sorting, which actually reduces the predicted exposure of high-income Black individuals relative to if they uniformly visited venues belonging to the chains they frequent in the CBSAs where they reside.

Why do high-income Black individuals choose lower income venues than predicted based on cost of travel distance alone, despite preferring high-income co-patrons? Our model rationalizes these choices through racial homophily combined with a strong correlation between the racial and income compositions of venues. Given that heavily Black (or Hispanic) venues often have lower income co-patrons, Black (or Hispanic) individuals face a trade-off between visiting heavily high-income venues and visiting heavily own-race venues that White (or

[31] Chain choice depends on taste for chain product offering, preferences for average co-patron composition within that chain, and distance to venues within that chain. Given the small importance of chain choice, however, we conclude that most of social exposure is determined by venue choice within, not across, chains.

Asian) individuals do not face.[32] This trade-off is quantitatively important. If we set preferences for own-race shares to zero but keep income preferences at their estimated values, the predicted income exposure of high-income Black becomes two percentage points *higher* than that of White individuals. Further, high-income Black individuals also live close to venues with more Black co-patrons, but fewer high-income co-patrons than the average venue.[33] As suggested by Bayer and McMillan (2005), the scarcity of high-income majority Black neighborhoods may explain why Black households live in poorer neighborhoods than White households with similar incomes.

Preferences are aligned with neighborhood sorting among other high-income groups. For high-income Asian, Hispanic and White individuals, either neighborhood sorting or preferences can explain most of the income exposure gap between uniform visits within CBSA/chain and actual visits. In other words, high-income Asian, Hispanic, and White individuals live in neighborhoods where venues that suit their social preferences, with large shares of high-income co-patrons, are located nearby. These results are consistent with preferences for social exposure playing an important role in neighborhood choice. The next section investigates this further.

6.2 Sorting on social preferences

We now use our model of venue choice to study the extent of sorting on social preferences and begin to understand how this sorting comes about.[34] Rather than estimating preferences over all neighborhood attributes, we study the evolution of preferences over co-patron composition around residential moves. We estimate the model for individuals in the same demographic group that reside in and move between neighborhoods with different demographic composition. Individuals in the same demographic group are sorted across neighborhoods in a way that is correlated with their preferences for social exposure. When individuals move, they move to neighborhoods that suit their different ex-ante social preferences and these differences grow gradually in the 6 months after the move.

[32] As mentioned in Section 3 Asian individuals have few heavily own-race venues regardless of incomes, and White individuals have heavily own-race venues that tend to be heavily high-income.

[33] Appendix Table C.3 shows that this neighborhood sorting explains a marked shift in the predicted own-race exposure of high-income Black individuals, from 0.15 based on chain and CBSA choice alone, but increases to 0.27 when accounting for neighborhood sorting.

[34] Persistent racial segregation of residences has been attributed to wealth differences, preferences, prejudice, and housing-market discrimination (Charles, 2003; Rothstein, 2017).

6.2.1 The extent of sorting on social preferences

We first examine whether people living in higher-income or heavily own-race neighborhoods have stronger preferences for higher-income or heavily own-race co-patrons. We divide Census tracts into (population-weighted) terciles by the share of residents who are high-income or in each racial/ethnic demographic. We re-estimate our baseline model from equation (2) using the devices that live in each tercile of the resident income and own-race distributions. Figure 5 depicts these preference estimates for high- and low-income White individuals (solid and dashed lines, respectively).[35]

Within a demographic group, we find that preferences are aligned with neighborhood demographic mix. The left-hand plot in Figure 5 Panel A shows that White individuals who reside in higher-income neighborhoods have stronger preferences for high-income co-patrons. This is true for both high- and low-income residents. In fact, the differences in preferences across terciles of neighborhood income are the same magnitude as the differences in preferences between high- and low-income individuals: a low-income resident of a top-income tercile neighborhood (dashed red) has similar preferences for co-patron income as a high-income resident of a middle-income tercile neighborhood (solid gray).[36] The right-hand plot of Panel B shows a similar alignment in own-race preferences with residential neighborhood demographics: White individuals who reside in neighborhoods with more White residents exhibit stronger preferences for own-race co-patrons. This difference is particularly pronounced between the bottom and upper-two terciles of own-race neighborhoods. Those in the top neighborhood tercile would travel five times farther than those in the bottom tercile to visit a venue in the top decile of White co-patron share rather than the bottom decile.

By contrast, we find almost no differences in preferences for co-patron income across residents in neighborhoods with different own-race levels or, conversely, no differences in preferences for co-patron own-race share across residents in neighborhoods with different income levels.

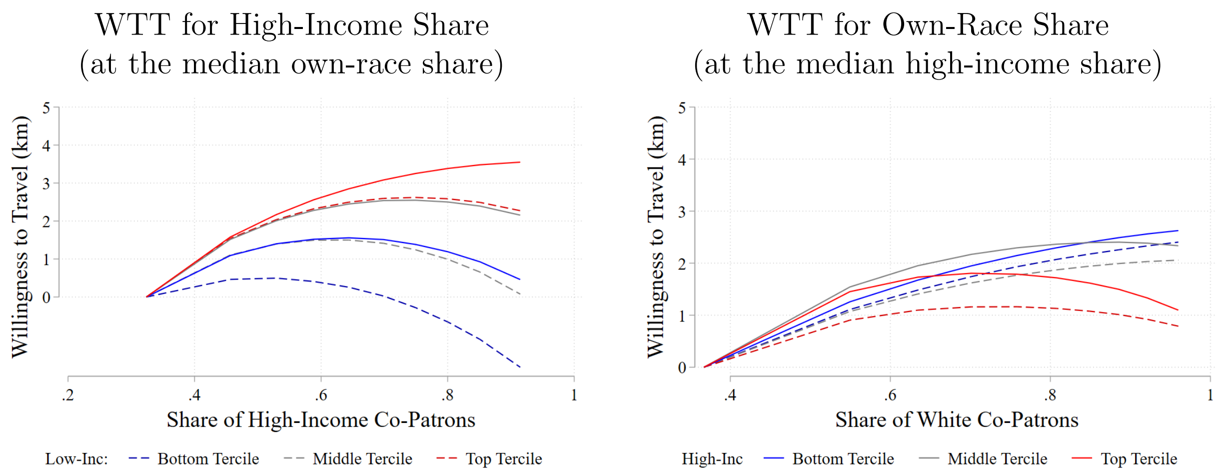
Within race-income groups, residential demographics and preferences over co-patron demographics are meaningfully correlated. Low-income individuals who reside in high-income neighborhoods much prefer high-income co-patrons, and White individuals who reside in

[35] Analogous figures for other racial/ethnic groups are in Appendix C.6. They show generally similar patterns, but are noisier due to smaller samples. For example, few low-income Black individuals live in the upper terciles of the tract income distribution.

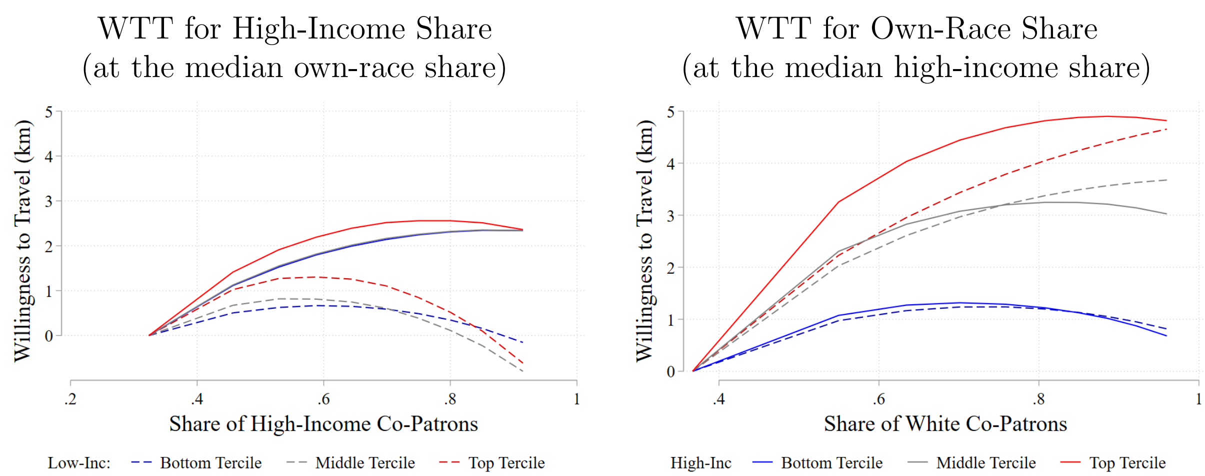
[36] One may worry that spatial heterogeneity in preferences for high-income may be due to, for instance, the group we define as low-income having higher incomes in the upper tercile of the tract income distribution. We note that the spatial heterogeneity we find is so large that our two income-group breakdown is enough to unambigiously exclude these differences explaining all the observed spatial heterogeneity. For instance, the willingness to travel of low-income White devices in the upper tract income tercile is higher than the willingness to travel of high-income White devices in the bottom income tercile.

Figure 5: Preference heterogeneity for White visitors

Panel A: Residential Tract High-Income Tercile



Panel B: Residential Tract Own-Race Tercile



Notes: The figures are analogous to Figure 4. Residential tract high-income terciles are defined using tract-level high-income share weighted by high-income tract population. Similarly for own-race terciles. In addition to our regular device selection criteria, each estimation sample contains only devices that live in tracts in a given residential tercile. We draw 100,000 visit cases for each spatial heterogeneity group. To compare across terciles, we evaluate willingness to travel relative to the same fixed demographic composition in all terciles. On the left, we show willingness to travel relative to a venue at the bottom decile of the high-income share distribution across all venues, holding own-race share fixed at its median value. On the right, we show willingness to travel relative to a venue at the bottom decile of the own-race share distribution across all venues, holding high-income share fixed at its median value.

the most White neighborhoods have stronger preferences for same-race co-patrons. These patterns could arise because individuals have similar preferences over their neighbors' and their co-patrons' demographics or through mechanisms that link residential experiences to preferences over co-patrons, such as the intergroup contact hypothesis.

6.2.2 Origin of Social-Preference Sorting Patterns

To examine the connections between residential demographics and preferences over co-patron demographics, we estimate these preferences before and after a move to a new city. To achieve sufficient sample sizes, we estimate monthly preference parameters of high-income White individuals who move between CBSAs for five months before and five months after their change in residence.[37] We examine how preference parameters evolve with moves across the neighborhood own-race distribution.[38] We estimate the parsimonious linear version of our model—used for all robustness exercises in Section 5—in which preferences over co-patron composition depend on only the share of own-race co-patrons and share of high-income co-patrons. We allow both transit costs (δ^g) and preferences over co-patron composition (β^g) to vary month by month:

$$Y_{ijt} = \sum_{k=-5}^{5} \mathbf{1}\{t = k\} \left(\delta_{1k}^{g,od} \ln \text{distance}_{ijt} + \delta_{2k}^{g,od} \ln \text{distance}_{ijt}^2 + \beta_{rk}^{g,od} \text{ownrace}_j + \beta_{yk}^{g,od} \text{highinc}_j\right) \tag{7}$$

Table 4: Average Own-Race Preference of High-Income White Devices Before & After Moves

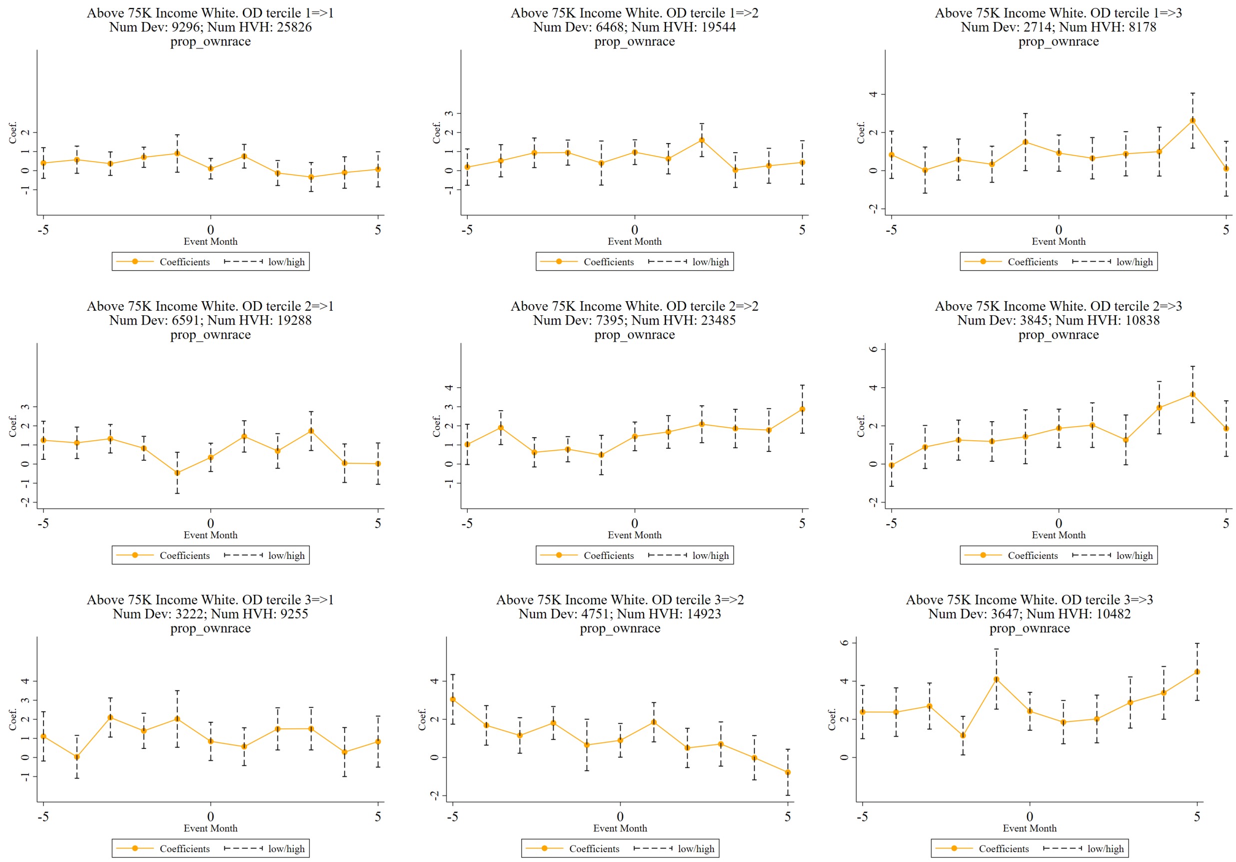
	O1			O2			O3		
	D1	D2	D3	D1	D2	D3	D1	D2	D3
Pre-Move	0.59	0.60	0.65	0.81	0.96	0.94	1.33	1.67	2.54
	(0.83)	(1.00)	(1.35)	(0.97)	(1.00)	(1.29)	(1.33)	(1.25)	(1.46)
Post-Move	0.06	0.65	1.03	0.71	1.96	2.27	0.92	0.53	2.84
	(0.90)	(1.09)	(1.52)	(1.15)	(1.23)	(1.60)	(1.40)	(1.33)	(1.56)

NOTES: We sample home-venue-home visits to restaurants by only cross-CBSA high income White movers and split movers by origin-destination own race tercile pairs.

[37] We focus on between-CBSA moves so that there is a stark change in the choice set and no scope for venue-specific habits to drive behavior. We follow devices for only ten months because most devices appear in the data for less than a year. The estimation sample is an unbalanced panel: not every mover is in the sample for all ten months and not every mover makes a home-venue-home restaurant visit in every month. High-income White individuals are by far the largest sample of cross-CBSA movers.

[38] Appendix C.2 reports results for moves across the neighborhood income distribution.

Figure 6: Movers Result Across Own Race Terciles: Own Race Preference



NOTES: Each plot in this figure depicts event-month-specific, origin-destination-race-tercile-specific coefficient estimates of $\beta^{g,od}_{k,r}$ from equation (7). Each point represents the coefficient on the own race co-patron share for a given month since moves occurred, with the bands reflecting 95% confidence intervals on those estimates. We sample home-venue-home visits to restaurants by only cross-CBSA high-income White movers and split movers by origin-destination own race tercile pairs.

Figure 6.2.2 shows estimated preferences for own-race exposure before and after moves across terciles of the neighborhood own-race distribution.[39] Given the small samples of movers, the month-specific parameter estimates are noisy. Table 4 reports the pre-move and post-move five-month averages of these coefficients for each event study.

The results reported in Figure 6.2.2 and Table 4 demonstrate three patterns of interest:

[39] Appendix C.2 includes three analogous sets of plots depicting estimates of preferences for income exposure before and after a move across neighborhoods in different own-race tercile pairs, and of preferences for own-race and income exposure following moves across neighborhoods in different income tercile pairs (with the limitation mentioned above). Out of 36 event studies, we note one unexplained jump in preferences following moves from the highest tercile to the bottom tercile of the income distribution. These movers to substantially poorer neighborhoods appear to experience an immediate drop in their preferences for high-income exposure upon moving. This result is difficult to interpret because such movers are rare and may have experienced a negative income shock.

1. Individuals move into neighborhoods whose demographic mix suits their pre-move social preferences. We find that pre-move racial preferences are generally stronger for destinations neighborhoods with higher white shares, conditional on the demographic mix of the origin. These differences are not always statistically significant, but they are especially large – a near doubling in the strength of racial homophily – when comparing someone moving from O3 to D3 with someone moving from O3 to D1. In other words, someone moving from a high to a low share white neighborhood has subtantially weaker racial preferences to begin with.

2. Estimated preference coefficients do not jump discontinuously when individuals move. There are nine different event studies in Figure 6.2.2, for each possible combination of moves across terciles of the own-race neighborhood distribution. The top-right and bottom-left plots show the widest moves possible in the own-race space, i.e., from an origin in the bottom tercile (O1) to a destination in the top tercile (D3) of the white share distribution, and from O3 to D1. In all cases, there are no pre-trends in preferences prior to a move, or noticeable jump in preferences right after moving.

3. Social preferences (slowly) converge to the local demographic mix after a move. The difference between the pre- and post-move social preference estimates are consistent with the idea that neighborhoods might, at least to some extent, shape preferences. For instance, someone moving from O1 to D3 experiences stronger racial preferences post-move relative to prior to a move, but still much weaker than the racial preferences of a mover whose origin was already within the top tercile of white share (O3 to D3). Given noisy estimates and a limited time period, we conclude that while preferences likely evolve following a move, social preferences do not appear to rapidly converge to those of incumbent residents.[40]

If preferences prior to a move predict the dominant demographic of the destination neighborhood, then sorting might be important. If preferences after a move evolve towards stronger preferences for the dominant demographic of the destination neighborhood, then repeated contact in adulthood may play a role in generating our preference estimates.

Overall, our movers design suggests that spatial sorting may underlie the wide spatial variation in preferences for social exposure that we document. In turn, social exposure to new neighbors may slowly shape individual preferences. Our results, for instance, would be

[40]The evidence for sorting towards preferred neighborhood and for neighborhoods shaping preferences is not always significant (albeit in the right direction). That said, comparing post-move preferences across destination terciles, we see large and significant or nearly significant evidence for the combinations of both effects (sorting and shaping) within all three origin terciles.

consistent with extended contact with one's neighbors early in life determining preferences for social exposure, which then evolve only slowly through one's adult life. Despite their current limitations, our movers results demonstrate the potential for mobility data to spur considerable progress in our understanding of the origin of social preferences, and of the determinants of social interactions.

7 Conclusion

Americans' daily activities are fragmented along demographic and economic lines. In this paper, we use smartphone movement data to estimate individual preferences over the demographic and economic composition of co-patrons in commercial venues. We then investigate determinants of the gaps between experienced and preferred social exposure for different demographic groups.

We do not find systematic differences in preferences for high-income co-patrons between racial/ethnic groups. But symmetric preferences do not yield symmetric outcomes: Black and Hispanic individuals visit establishments with high-income co-patrons less often than White individuals. Black and Hispanic preferences for social exposure to high-income individuals go unsatisfied for two reasons: racial differences in proximity to high-income venues and racial differences in the correlation between high-income share and own-race share. Establishments with many Black and Hispanic co-patrons tend to have fewer high-income co-patrons, in part, because Black and Hispanic people are more likely to be low-income. To the extent that exposure to high-income co-patrons is important for economic mobility, residential segregation and racial homophily reinforce racial differences in incomes.

The preferences estimated using choices within restaurant chains are linked to demographic differences in other economic domains. For example, we document that White individuals residing in heavily White neighborhoods prefer same-race co-patrons more than White individuals residing in more racially diverse neighborhoods. This alignment of neighborhood demographics and preferences over co-patron demographics suggests an alignment of preferences over both neighbors' and co-patrons' demographics or mechanisms that link preferences over co-patrons to residential experiences.

References

Abbiasov, Timur, Cate Heine, Edward L Glaeser, Carlo Ratti, Sadegh Sabouri, Arianna Salazar Miranda, and Paolo Santi. 2022. "The 15-Minute City Quantified Using Mobility Data." National Bureau of Economic Research.

Akbar, Prottoy A., Victor Couture, Gilles Duranton, and Adam Storeygard. 2023. "The Fast, the Slow, and the Congestion: Urban Transportation in Rich and Poor Countries." Processed, University of British Columbia.

Aliprantis, Dionissi, Daniel R. Carroll, and Eric R. Young. 2022. "What explains neighborhood sorting by income and race?" *Journal of Urban Economics*, 103508.

Anderson, Elijah. 2011. *The Cosmopolitan Canopy: Race and Civility in Everyday Life.* W.W. Norton & Company.

Athey, Susan, Billy Ferguson, Matthew Gentzkow, and Tobias Schmidt. 2021. "Estimating experienced racial segregation in US cities using large-scale GPS data." *Proceedings of the National Academy of Sciences*, 118(46): e2026160118.

Athey, Susan, David Blei, Robert Donnelly, Francisco Ruiz, and Tobias Schmidt. 2018. "Estimating Heterogeneous Consumer Preferences for Restaurants and Travel Time Using Mobile Location Data." *AEA Papers and Proceedings*, 108: 64–67.

Atkin, David, Keith Chen, and Anton Popov. 2022. "The Returns to Face-to-Face Interactions: Knowledge Spillovers in Silicon Valley."

Baldenius, Till, Nicolas Koch, Hannah Klauber, and Nadja Klein. 2023. "Heat increases experienced racial segregation in the United States." arXiv.org Papers 2306.13772.

Barwick, Panle Jia, Yanyan Liu, Eleonora Patacchini, and Qi Wu. 2023. "Information, Mobile Communication, and Referral Effects." *American Economic Review*, 113(5): 1170–1207.

Bayer, Patrick, and Robert McMillan. 2005. "Racial Sorting and Neighborhood Quality." NBER working paper.

Bayer, Patrick, Hanming Fang, and Robert McMillan. 2014. "Separate when equal? Racial inequality and residential segregation." *Journal of Urban Economics*, 82: 32–48.

Bayer, Patrick, Marcus D Casey, W. Ben McCartney, John Orellana-Li, and Calvin S Zhang. 2022. "Distinguishing Causes of Neighborhood Racial Change: A Nearest Neighbor Design." National Bureau of Economic Research Working Paper 30487.

Bayer, Patrick, Stephen L. Ross, and Giorgio Topa. 2008. "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes." *Journal of Political Economy*, 116(6): 1150–1196.

Black, Sandra. 1999. "Do Better Schools Matter? Parental Valuation of Elementary Education." *The Quarterly Journal of Economics*, 114(2): 577–599.

Büchel, Konstantin, Maximilian V Ehrlich, Diego Puga, and Elisabet Viladecans-Marsal. 2020. "Calling from the outside: The role of networks in residential mobility." *Journal of Urban Economics*, 119: 103277.

Caetano, Gregorio, and Vikram Maheshri. 2019. "Gender segregation within neighborhoods." *Regional Science and Urban Economics*, 77: 253–263.

Caetano, Gregorio, and Vikram Maheshri. 2021. "A Unified Empirical Framework to Study Segregation."

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

Charles, Camille Zubrinsky. 2003. "The Dynamics of Racial Residential Segregation." *Annual Review of Sociology*, 29(1): 167–207.

Chetty, Raj, John N Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan. 2020. "Income Segregation and Intergenerational Mobility Across Colleges in the United States." *The Quarterly Journal of Economics*, 135(3): 1567–1633.

Chetty, Raj, Matthew O Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, et al. 2022*a*. "Social capital I: measurement and associations with economic mobility." *Nature*, 608(7921): 108–121.

Chetty, Raj, Matthew O Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, et al. 2022*b*. "Social capital II: determinants of economic connectedness." *Nature*, 608(7921): 122–134.

Combs, James G, and David J Ketchen Jr. 2003. "Why do firms use franchising as an entrepreneurial strategy?: A meta-analysis." *Journal of management*, 29(3): 443–465.

Cook, Cody. 2023. "Heterogeneous Preferences for Neighborhood Amenities: Evidence from GPS data." *Available at SSRN 4212524*.

Cook, Cody, Lindsey Currier, and Edward L Glaeser. 2022. "Urban Mobility and the Experienced Isolation of Students." National Bureau of Economic Research Working Paper 29645.

Couture, Victor, Cecile Gaubert, Jessie Handbury, and Erik Hurst. 2023. "Income Growth and the Distributional Effects of Urban Spatial Sorting." *The Review of Economic Studies*, rdad048.

Couture, Victor, Gilles Duranton, and Matthew A. Turner. 2018. "Speed." *Review of Economics and Statistics*, 100(4): 725–739.

Couture, Victor, Jonathan I Dingel, Allison Green, Jessie Handbury, and Kevin R Williams. 2021. "JUE Insight: Measuring movement and social contact with smartphone data: a real-time application to COVID-19." *Journal of Urban Economics*, 103328.

Davis, Donald R., Jonathan I. Dingel, Joan Monras, and Eduardo Morales. 2019. "How Segregated Is Urban Consumption?" *Journal of Political Economy*, 127(4): 1684–1738.

Davis, Morris A., Jesse Gregory, and Daniel A. Hartley. 2023. "Preferences over the Racial Composition of Neighborhoods: Estimates and Implications."

Echenique, Federico, and Roland G. Jr. Fryer. 2007. "A Measure of Segregation Based on Social Interactions." *The Quarterly Journal of Economics*, 122(2): 441–485.

Goldszmidt, Ariel, John A List, Robert D Metcalfe, Ian Muir, V. Kerry Smith, and Jenny Wang. 2020. "The Value of Time in the United States: Estimates from Nationwide Natural Field Experiments." National Bureau of Economic Research Working Paper 28208.

Klopack, Ben. 2020. "One size fits all? The value of standardized retail chains." Texas A&M University working paper.

Krueger, Alan B. 1991. "Ownership, Agency, and Wages: An Examination of Franchising in the Fast Food Industry." *The Quarterly Journal of Economics*, 106(1): 75–101.

Massenkoff, Maxim, and Nathan Wilmers. 2023. "Rubbing Shoulders: Class Segregation in Daily Activities." https://ssrn.com/abstract=4516850.

McFadden, Daniel. 1978. "Modelling the Choice of Residential Location." In *Spatial Interaction Theory and Planning Models.* , ed. A. Karlqvist, L. Lundqvist, F. Snickars and J. Weibull, 75–96. Amsterdam:North Holland.

Miyauchi, Yuhei, Kentaro Nakajima, and Stephen J. Redding. 2021. "Consumption Access and Agglomeration: Evidence from Smartphone Data." CEPR Discussion Paper 6741-1613580728.

Moro, Esteban, Dan Calacci, Xiaowen Dong, and Alex Pentland. 2021. "Mobility patterns are associated with experienced income segregation in large US cities." *Nature communications*, 12(1): 4633.

Reardon, Sean F., Lindsay Fox, and Joseph Townsend. 2015. "Neighborhood Income Composition by Household Race and Income, 1990–2009." *The ANNALS of the American Academy of Political and Social Science*, 660(1): 78–97.

Rothstein, Richard. 2017. *The Color of Law: A Forgotten History of How Our Government Segregated America.* Liveright.

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

Schönholzer, David. 2023. "Measuring Preferences for Local Governments."

Sethi, Rajiv, and Rohini Somanathan. 2004. "Inequality and Segregation." *Journal of Political Economy*, 112(6): 1296–1321.

Small, Kenneth A, and Erik T Verhoef. 2007. *The economics of urban transportation.* Routledge.

Train, Kenneth. 2009. *Discrete Choice Methods with Simulation.* Cambridge University Press.

Waldfogel, Joel. 2008. "The median voter and the median consumer: Local private goods and population composition." *Journal of Urban Economics*, 63(2): 567–582.

Wang, Qi, Nolan Edward Phillips, Mario L. Small, and Robert J. Sampson. 2018. "Urban mobility and neighborhood isolation in America's 50 largest cities." *Proceedings of the National Academy of Sciences*, 115(30): 7735–7740.

Williamson, Oliver E. 1991. "Comparative Economic Organization: The Analysis of Discrete Structural Alternatives." *Administrative Science Quarterly*, 36(2): 269–296.

Xu, Yang, Alexander Belyi, Paolo Santi, and Carlo Ratti. 2019. "Quantifying segregation in an integrated urban physical-social space." *Journal of The Royal Society Interface*, 16(160): 20190536.

Appendix – For Online Publication

A Data Appendix

A.1 Smartphone visits data

As described in Couture et al. (2021), each observed visit consists of a device, a venue, a timestamp, and an attribution score. PlaceIQ's attribution scores are larger when a device is more likely to have been within a venue, based on the number and density of pings, data source of pings, and proximity of the pings to the polygon defining the venue. We retain all visits with an attribution score greater than a threshold value recommended by PlaceIQ based on their experience correlating their data to a diverse array of truth sets, including consumer spending data and foot-traffic counts. PlaceIQ also reports a lower bound for the visit's duration based on the time between consecutive pings at the same venue.

We also clean the visit data to remove simultaneous visits. For instance, when two venues are in close proximity to one other, a single visit event may have an attribution score for both venues that exceeds the threshold value recommended by PlaceIQ. We retain only the visit to the venue with the highest attribution score. In other cases, the polygons of two different venues overlap.[41] When two polygons overlap, we retain polygons with an identified business category over those lacking a category.

Table A.1 summarizes the smartphone movement data after this cleaning for days between June 1, 2018 and December 31, 2019. On the average day, there were 167 million visits produced by 38 million devices visiting 40 million residential and non-residential venues. The average device appears in the data for 159 days over the 19 month window, but a notable number appear on only one day. After we restrict attention to devices in our estimation sample (one permanent home assignment over 19 month window) there are 104 million visits from 18 million devices visiting 30 million venues on an average day.

A.2 Home assignments

We construct home assignments using a procedure introduced in Couture et al. (2021), which we repeat here for convenience. Residential venues are a distinct category in the PlaceIQ data. This allows us to construct a weekly panel of home locations for a subset of devices using the following assignment methodology:

[41] This could happen, for instance, if the basemap contains one polygon representing a business establishment and a second polygon representing both that building and the accompanying parking lot.

Table A.1: Summary statistics for cleaned visits

	Cleaned visits sample			
	Mean	SD	5th	95th
Devices	38.18	4.68	30.4	44.61
Venues	39.55	2.77	34.38	42.97
Visits	166.51	18.17	137.32	193.3
Duration	159.24	172.83	1.0	524.0

NOTES: This table summarizes PlaceIQ data for June 1, 2018 to December 31, 2019 after our cleaning of the visits as described in the text. The counts of devices, venues, and visits are stated in millions per day. Duration is the number of days between a device's first and last appearance in the data (between June 1, 2018 and December 31, 2019).

1. For each week, we assign a device to the residential venue where its total weekly visit duration at night (between 5pm and 9am) is longest, conditional on it making at least three nighttime visits to that venue within the week.[42] If a device does not visit any residential location on at least three nights, then on initial assignment that device-week pair has a missing residential location.

2. After this preliminary assignment, we fill in missing weeks and adjust for noisiness in the initial panel using the following interpolation rules:

Rule 1: *Change "X · X" to "X X X"*: If the residential assignment for a week is missing and the non-missing residential assignment in the weeks before and after is the same, we replace the missing value with that residential assignment.

Rule 2: *"a X Y X b" to "a X X X b" where* $a \neq Y$ *and* $b \neq Y$: If a device has a residential assignment Y that does not match the assignment X in the week before or after, we replace Y with X as long as Y was not the residential assignment two weeks before or two weeks after.[43]

3. After step 2's interpolation, for any spells of at least four consecutive weeks where a device is assigned the same residential venue, we assign that venue as a device's "home" for those weeks. Spells of less than four weeks are set to missing.

4. If a device has more than one home assignment and the pairwise distance between them is less than 0.1 kilometers, we keep the home that appears for the most weeks.

[42] Since we only observe minimum duration, there are instances where total duration is 0 across all residential locations. In these cases, we assign the residential venue as the venue a device makes the most nighttime visits.

[43] For cases where a device's residential location is bouncing between two places ("Y X Y X X") we are not able to ascertain whether Y or X is more likely to be a device's residence in a given week

Table A.2: Homogeneous Buildings by Race and Income

Category	Group	Buildings	Percent
Race/Ethnicity	Buildings	34,857,456	100
	White	25,289,515	73
	Black	2,154,171	6
	Asian	957,125	3
	Hispanic	3726616	11
Income	Buildings	36,233,831	100
	Low Income	8,532,738	24
	Middle Income	13,071,836	36
	High Income	14,145,344	39

NOTES: This table shows the number of buildings for we have information on race/ethnicity and income. The table also shows the number of buildings that are "nearly homogeneous" ($> 67\%$) for the four race/ethnicity groups and three income groups.

5. If a device has the same home assignment in two non-consecutive periods and no other home assignments in between, then we assign all weeks in between to that home assignment.

A.3 Building-level Demographics

PlaceIQ provides us with demographic data at the building level for around 36 million residential buildings. This includes information on standard demographic information such as education, income, race, gender, and age. Each category is reported in discretized buckets, and a building is assigned weights across buckets reflecting the share of people who live in the building who fall into each bucket. For income, we aggregate the provided bins to low-income ($< \$50,000$), middle-income ($\$50,000 - \$100,000$), and high-income ($\$100,000+$). For racial/ethnic categories, we aggregate the provided bins to non-Hispanic Black, non-Hispanic white, non-Hispanic Asian, Hispanic, and all other racial/ethnic groups.

Table A.2 shows the number of buildings that contain information on race/ethnicity and income. In the underlying data, some buildings only contain information for certain demographic categories. This is reflected in the 1.5 million additional buildings that have income data but lack data on race/ethnicity. The table also shows the number of buildings that are "nearly homogeneous" ($> 67\%$) for the four race/ethnicity groups and three income groups. 99% of buildings are at least 67% low-, middle-, or high-income. 93% of buildings are at least 67% white, Black, Hispanic, or Asian.

A.4 Building-level data quality and representativeness

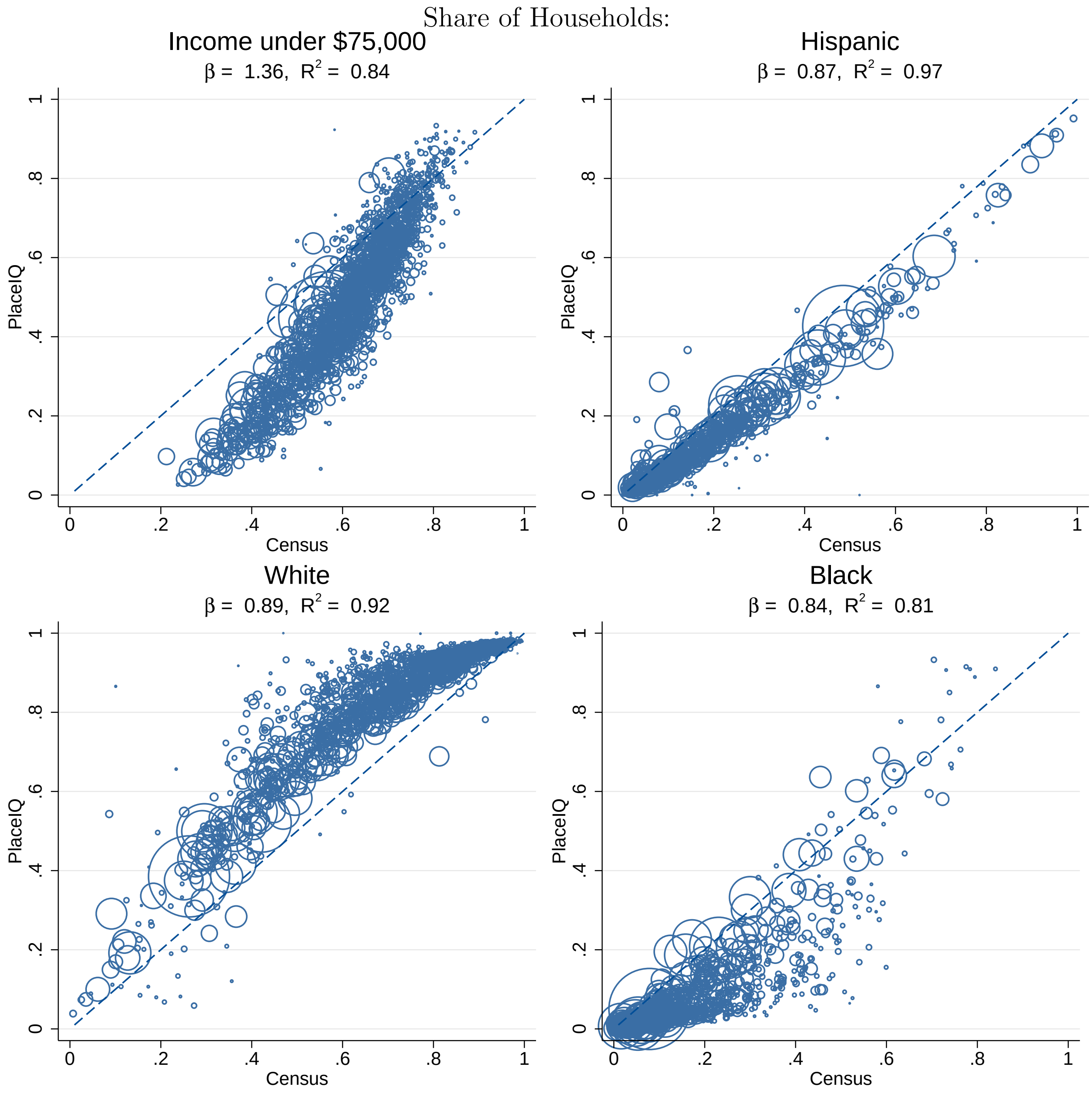
The building-level demographic data are highly correlated with publicly available Census demographic data when aggregated to larger spatial units. Figure A.1 compares four county-level demographic shares in the building-level data to those in the 2015-2019 American Community Survey (ACS): share of non-Hispanic Black residents, share of non-Hispanic white residents, share of Hispanic residents, and share of residents whose household income is less than $75,000.[44] The two county-level measures are highly correlated: the R^2 exceeds 0.80 for all four demographic shares. Couture et al. (2021) find that device populations are proportionately distributed across block groups within counties, so the gaps between observations and the 45-degree line in Figure A.1 largely reflect differences within, rather than across, block groups.

Aggregating the building-level demographic information to counties yields more White and high-income households than found in the Census data. Figure A.1 shows that, on average, the population share of White residents is 28 percentage points higher in the PlaceIQ data than in the Census data, and the population share of high-income households is 32 percentage points higher. These differences also vary in intensity across counties. The top-left plot of Figure A.1 shows that the share of low-income devices in the PlaceIQ building-level data is smaller than the share of low-income residents in the Census, except in counties with the largest low-income shares. This means that low-income households are under-represented in the PlaceIQ data, but less so in counties with more low-income households. Finally, the three other plots of Figure A.1 show that, compared to the Census data, Hispanic households are proportionally represented while White households are over-represented and Black households are under-represented in the PlaceIQ data. These differences may reflect measurement error in the building-level data or non-uniform selection into the smartphone movement data. We also observe a higher device-to-population ratio in counties that have more white and high-income households, which is consistent with selection into the smartphone data.

We next compare patterns of visit behavior between buildings in the same block group, to establish that the building-level data is informative at a finer level than the most granular demographic data publicly available from the Census Bureau. Waldfogel (2008) and Klopack (2020) find that race and income correlate with heterogeneity in preferences for different types of venues and chains. We therefore expect building-level demographic differences to generate observable differences in visit patterns between people of different demographic groups living in the same block group (and therefore facing the same choice set). These observable differences in behavior predicted using only building-level demographic data, in

[44] The phrases "White residents" and "Black residents" henceforth indicate non-Hispanic White and non-Hispanic Black residents respectively.

Figure A.1: Comparison of county-level demographics in Census and building-level data



NOTES: These plots compare county-level demographic composition in the 2015-2019 American Community Survey and the building-level demographic data. The diameter of each marker is proportionate to the county's population in the ACS. The regression coefficient and R^2 reports the result of regressing PlaceIQ county shares on ACS shares weighted by the ACS population.

turn, should map into similar observable differences in behavior predicted using only demographic information from the Census. We therefore compare how across-building variation in demographics within a block group predicts chain popularity to how across-block-group variation in demographics within a tract predicts chain popularity. If demographics predict chain patronage, we should find similar rankings of chains using these two different data sources.

We proceed with this comparison as follows. First, we compute the within-block group chain popularity ranking, using building-level demographic information only. For each chain in each block group, we compute the ratio of the average number of visits by devices living in high-income ($> \$75,000$) buildings to the average number of visits by devices living in buildings that are not high-income.[45] We then take a weighted average of this ratio across block groups to obtain a ranking of the chains by popularity with high-income people relative to non-high income people.[46] Second, we compute the within-tract chain popularity ranking, using only Census block group demographic information. To do so, we compute, for each chain in each census tract, the average ratio of visits for devices living in block groups that are at least 67% high-income, to visits from devices living in block groups that are least 67% not high-income. Finally, we compute an analogous set of comparisons for ratios of visits by white versus non-white devices.

Figure A.2 depicts the results of these comparisons for restaurants and gas stations, the two business categories with the greatest number of chains. The plots in the first column show the relative propensity of high-income devices to visit a chain, and the plots in the second column shows the relative propensity of white devices to visit a chain. The ranking obtained using only building-level demographic variation within a block group is very similar to the ranking obtained using only Census demographic information. We find Spearman correlations between 0.7 and 0.9 for restaurant and gas chains, for both income and race. For example, both building-level and Census-block-group-level demographic information show

[45] We restrict attention to buildings that are homogeneous in the sense that at least 67% of their residents belong to one of these income groups.

[46] We weight using number of devices living in a block group (N_g) and a variance weight ($\sum_{i \in g} (r_{ig} - \bar{r}_g)^2$).

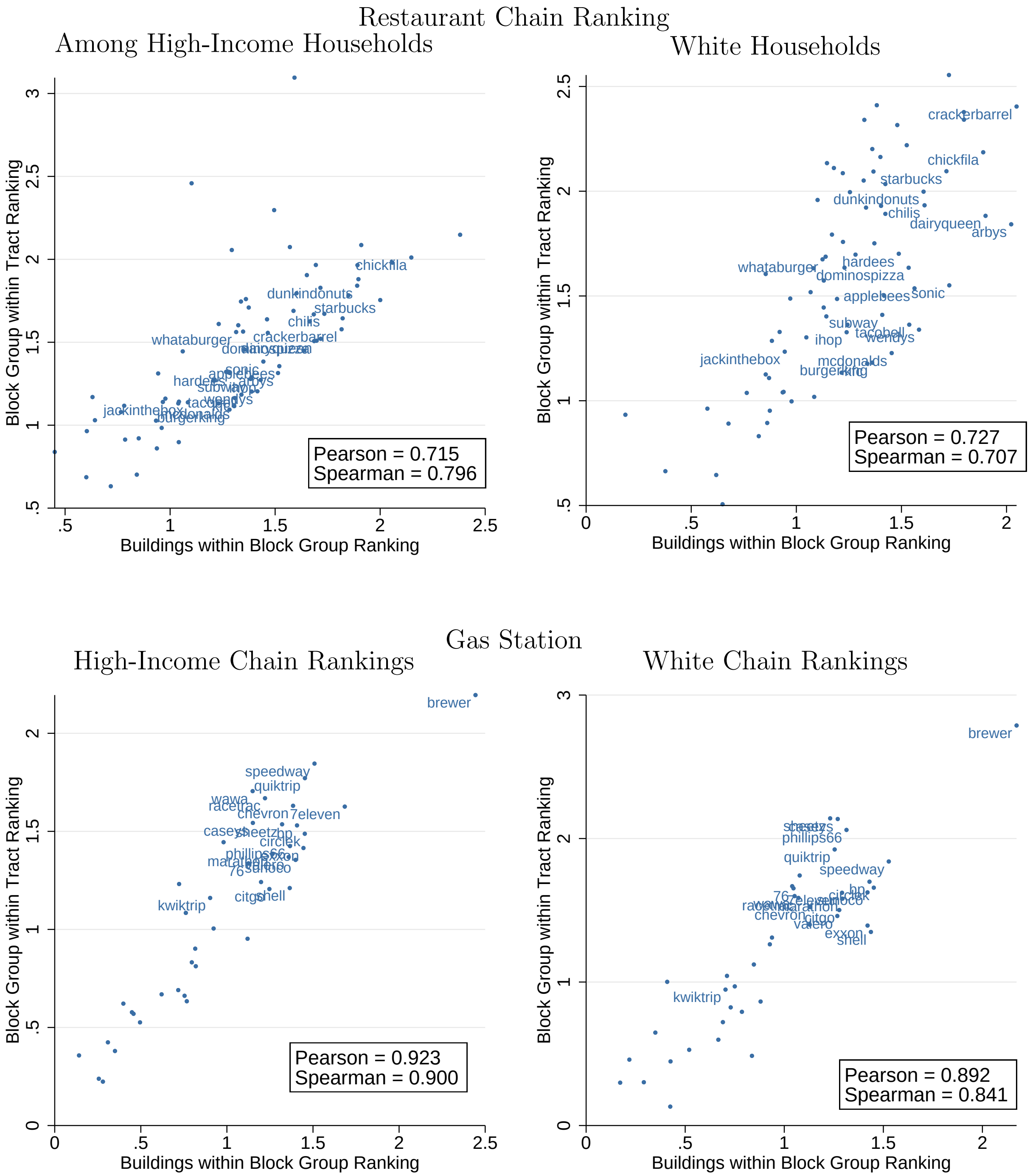
$$w_g = N_g \sum_{i \in g} (r_{ig} - \bar{r}_g)^2$$

Adding the second term produces a statistic that exactly matches the ranking of chains produced by the OLS estimate for γ_c for each chain:

$$\log y_{igc} = \gamma_c r_i z_c + \delta_{cg} z_c d_g$$

y_{igc} indicates the number of visits from device i living in block group g to chain c. r_i is an indicator if device i is of type r (high-income). z_c is an indicator if visit was to chain c. d_g is an indicator if device lives in block group g.

Figure A.2: Comparisons of chain popularity by demographic by spatial unit



NOTES: Each panel depicts the average ratio of visits to different chains by residents of different groups within a tract (vertical axis) or a block group (horizontal axis). Within-block-group variation is measured using buildings that are at least 90% one group in the building-level data. Within-tract variation is measured using block groups that are at least 67% one group in the 2015-2019 ACS. The first row shows chains in the "Restaurant" category, while the second row shows chains in the "Gas Station and Convenience Store" category. The first column compares the average ratio of visits by high-income residents ($> \$75,000$) to non-high-income residents, and the second column compares the average ratio visits by white residents to non-white residents. The largest 20 chains by number of visits within each category are labeled on each plot.

that high-income individuals make relatively more visits to Starbucks than Dunkin' Donuts.

A.5 Chain Coverage

Table A.3 summarizes information on the size of our venue sample for the five largest chain in each category of establishment (ten largest for restaurants). The table compares the actual number of establishments in each chain (gathered from various sources including company websites and investor reports) with the number of establishments in the PlaceIQ data. It also reports the total number of visits to each chain. The PlaceIQ basemap of venues is close to comprehensive, and contains upward 80 percent of all venues for most chains.[47] Bank and Gym chains receive fewer visitors than other categories, so preference estimates are noisier for these categories.

[47] For RiteAid and Walmart Neighborhood Market, we have more venues in the basemap than were open as of 3/31/2021. This reflects store closures rather than wrongly-identified locations.

Table A.3: Chain Venue Coverage in PlaceIQ Basemap

Category	Chain	Actual Count	PIQ Count	(%)	Number Visits (millions)
	Bank Of America		3124		0.68
	Wells Fargo		3035		0.66
Bank	Chase		3225		0.57
	PNC		1687		0.51
	Citizens Bank		823		0.34
	Walmart	4743	4429	93	393.25
	Target	1904	1360	71	66.89
Big Box	Costco	558	505	91	61.00
	SamsClub	599	555	93	47.54
	TractorSupplyCo	1923	1709	89	18.60
	Shell	12845	12635	98	280.94
	7Eleven	9364	7712	82	204.16
Gas	CircleK	7100	6061	85	163.07
	Exxon	11000	7633	69	153.78
	Chevron	7800	7742	99	147.02
	Kroger	2750	2247	82	103.73
	Safeway	1300	1315	101	35.56
Grocery	AholdDelhaize	2000	1524	76	31.50
	WalmartMarket	683	692	101	30.67
	Publix	1269	855	67	28.20
	Planet Fitness	1929	809		0.88
	LA Fitness		472		0.69
Gym	Orange Theory Fitness		373		0.54
	Anytime Fitness		837		0.53
	24 Hour Fitness		380		0.39
	CVSPharmacy	9900	8656	87	187.11
Pharmacies	Walgreens	9021	8380	93	135.51
	RiteAid	2500	2649	106	23.40
	Subway	22324	21693	97	992.49
	Starbucks	15328	10598	69	618.57
	McDonalds	13846	13050	94	580.77
	ChickFilA	2671	2030	76	224.12
	DunkinDonuts	8500	7719	91	218.47
Restaurant	BurgerKing	7257	6789	94	153.43
	Wendys	6500	5475	84	139.08
	TacoBell	6832	6743	99	138.03
	PizzaHut	6526	6085	93	129.12
	PandaExpress	2198	1630	74	101.41

Notes: This table shows the number of chain locations for the five largest chains in the venue categories of "Bank", "Big Box", "Gas", "Grocery", "Gym", "Pharmacies", and "Restaurant". The actual count reports the number of U.S. chain locations as reported by the company website or annual report to investors. The PlaceIQ count reports the total number of venues including those excluded from the estimation sample. The number of visits reports all visits to the chain between June 1, 2018 through December 31, 2019.

B Specification of mean utility

In the baseline specification, the component of utility that varies across venues within a chain, Y_{ij}, depends on the distance from the consumer's home to the venue (distance_{ij}), the share of high-income co-patrons (s_j^{highinc}), and the share of own-race co-patrons (s_j^{ownrace}):

$$Y_{ij} = \delta^g f_1(\ln \text{distance}_{ij}) + \beta^g f_2(s_j^{\text{ownrace}}, s_j^{\text{highinc}}),$$

where δ^g and β^g are group-specific coefficient vectors on transit costs and co-patron composition, respectively, $f_1(\ln \text{distance}_{ij})$ is a polynomial of log distance with unit coefficients and $f_2(s_j^{\text{ownrace}}, s_j^{\text{highinc}})$ is a polynomial of the two co-patron shares with unit coefficients.

Our baseline specification (2) uses second-degree polynomials:

$$\begin{aligned} Y_{ij} = {} & \delta_1^g \ln \text{distance}_{ij} + \delta_2^g \left(\ln \text{distance}_{ij}\right)^2 \\ & + \beta_1^g s_j^{\text{highinc}} + \beta_2^g \left(s_j^{\text{highinc}}\right)^2 + \beta_3^g s_j^{\text{ownrace}} + \beta_4^g \left(s_j^{\text{ownrace}}\right)^2 + \beta_5^g s_j^{\text{ownrace}} \times s_j^{\text{highinc}}. \end{aligned}$$

In this appendix, we first show the baseline parameterization $f_1(\ln \text{distance}_{ij})$ fits observed choice patterns better than a log-linear specification while performing similarly to higher order parameterizations. We then compare our baseline specification of co-patron prefereces, $\beta^g f_2(s_j^{\text{ownrace}}, s_j^{\text{highinc}})$, to an alternative where we discretize venues into clusters and show similar preference estimates. We find the baseline quadratic specification matches preferences estimates from the discretized comparison better than the linear specification while producing similar estimates to the cubic.

B.1 Co-Patron Composition

As an alternative representation of $f_2(s_j^{\text{ownrace}}, s_j^{\text{highinc}})$ we use K-means clustering to discretize venues into 50 groups.[48] We then run an alternative specification where mean utility is a function of distance and the cluster k to which venue j belongs.

$$Y_{ij} = \delta_1^g \ln \text{distance}_{ij} + \delta_2^g \left(\ln \text{distance}_{ij}\right)^2 + \sum_k \theta_k^g \mathbf{1}(j \in \mathcal{J}_k) \tag{B.1}$$

[48] The objective function for clustering venues by visitor income and race is

$$\underset{(S_1^g, S_2^g, \ldots, S_{50}^g)}{\arg\min} \sum_{j=1}^{50} \sum_{x \in S_j^g} \left\| x_j^g - \bar{x}_{S_j}^g \right\|^2,$$

where $x_j^g \equiv (\text{ownrace}_j^g, \text{highinc}_j^g$ and $\bar{x}_{S_j} \equiv \frac{\sum_{i \in S_j} x_i}{\sum_{i \in S_j} 1}$ is the mean value of x^g in S_j^g.

This parameterization of preferences over co-patrons is appealing because as the number of clusters gets very large this approximates to a non-parametric representation of preference over co-patrons.[49]

To compare the similarity of co-patron preference estimates in our baseline to the non-parametric approximation, we find the venue of race s_k^{ownrace} and income s_k^{highinc}) that is closest to the geometric mean of each cluster k. We then compute $\beta^g f_2(s_k^{\text{ownrace}}, s_k^{\text{highinc}})$ for each of the 50 clusters where f_2 can be any degree polynomial. By comparing, θ_k^g from Equation B.1 to this fitted value we benchmark the polynomial to a non-parametric approximation.

Figure B.1 shows the result of this comparison for the linear, quadratic, and cubic approximation. [50] Across all eight groups, the baseline quadratic specification is similar to the cluster estimates. The Pearson correlation exceeds 0.8 for every group except and exceeds 0.9 for seven of the eight groups. The cluster coefficients and quadratic fitted values disagree most for the least preferred venues. Given relatively monotonic preferences over race and income, the polynomial specification understates the extent visitors dislike venues with a small share of own-race and high-income co-patrons.

The baseline quadratic specification substantially outperforms the more parsimonious linear specification. Pearson correlations between the linear specifications and the non-parametric approximation vary from 0.14 (Low-Income Asian) to 0.76 (High-Income White) whereas for the baseline quadratic specification the correlations vary from 0.81 (Low-Income Asian) to 0.97 (High-Income Hispanic). However, adding additional terms for the cubic specification adds no additional predictive power. The cubic and quadratic specifications produce highly correlated preference estimates, and the Pearson correlation between cubic and quadratic fitted values exceed .97 for each of the eight groups.

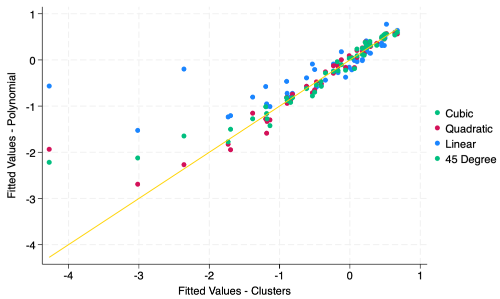
[49] There are two limitations to this approach such that we do not choose this for our baseline specification. First, we cannot compute the gradient for willingness to travel calculations (3). Second, measurement error is large in areas of the characteristic space where there are few venues.

[50] Fitted values are re-scaled to be mean zero.

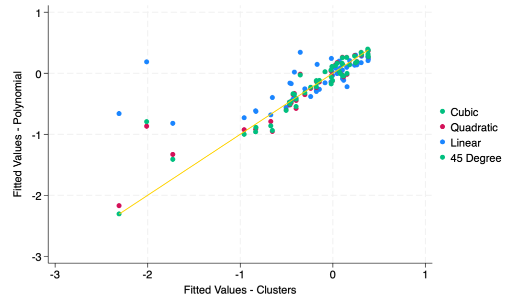
Figure B.1: Comparing specifications of $f_2(s_j^{\text{ownrace}}, s_j^{\text{highinc}})$

White

High-Income

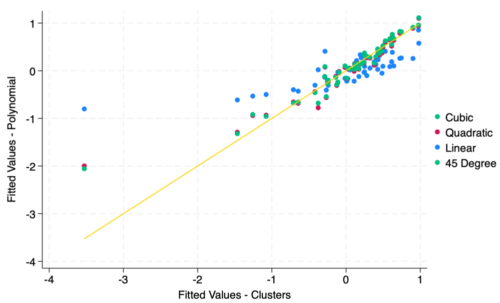


Low-Income

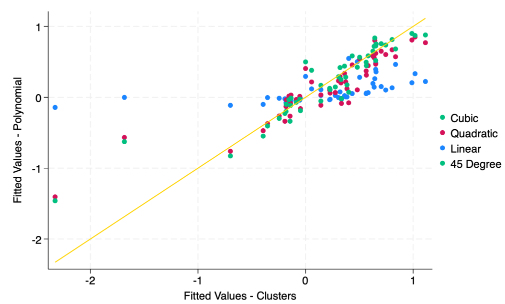


Black

High-Income

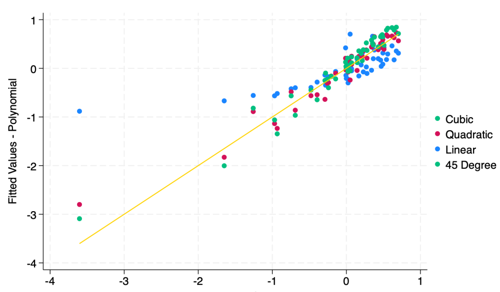


Low-Income

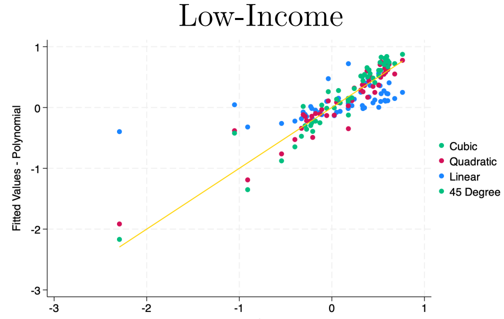


Hispanic

High-Income

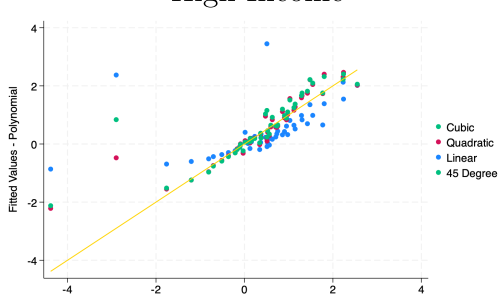


Low-Income

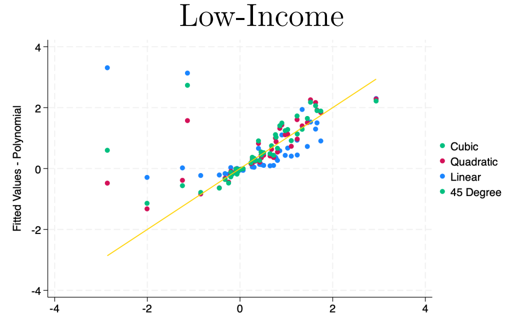


Asian

High-Income



Low-Income

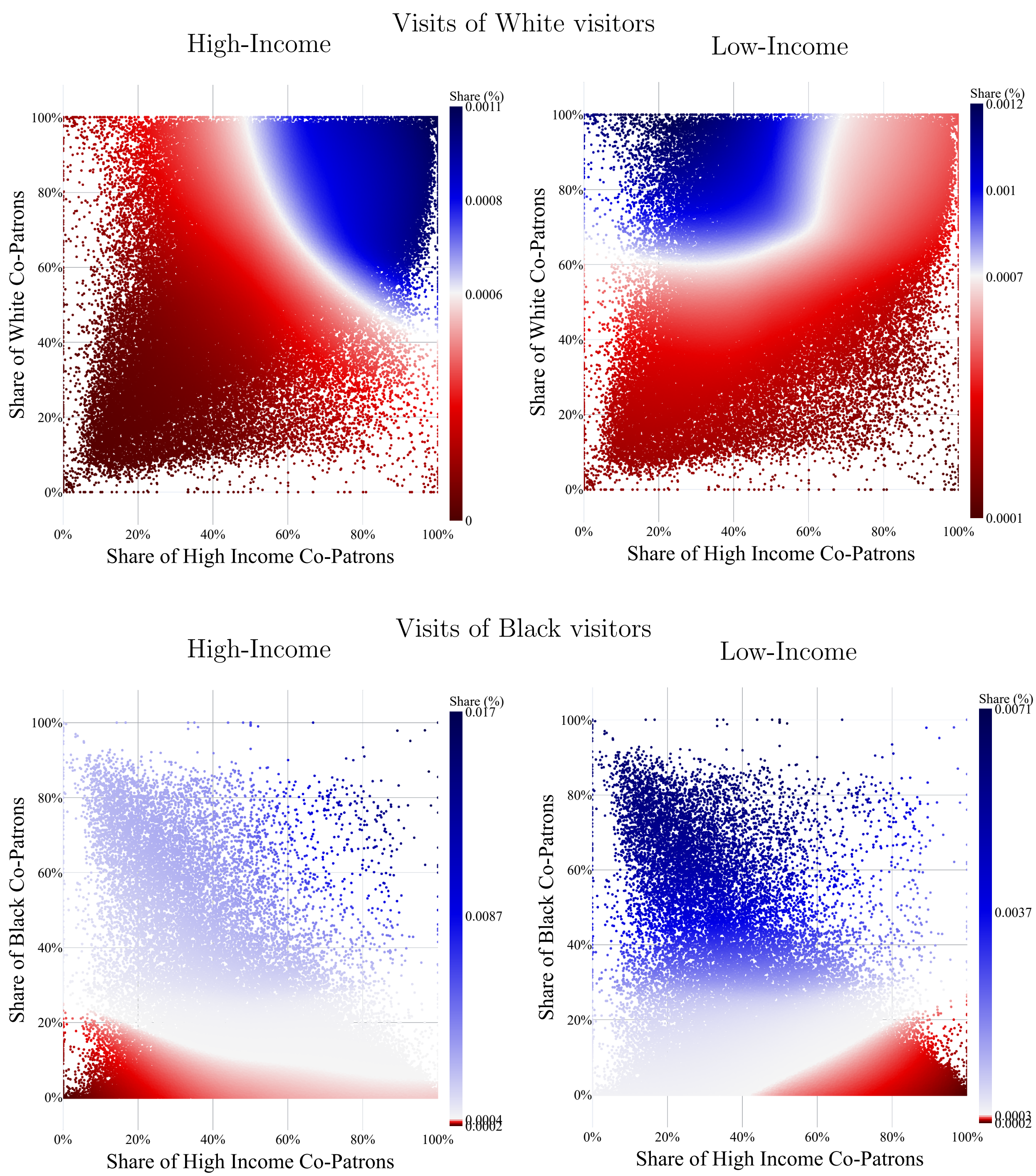


NOTES:

C Appendix Figures

C.1 Patterns of social exposure

Figure C.1: Exposure to (and Availability of) Co-Patron Mix in All Chains



Notes: Continued on next page. This figure is analogous to Figure 2 but shows venues in all business chain categories in our estimation sample.

Exposure to (and Availability of) Co-Patron Mix in All Chains cont.

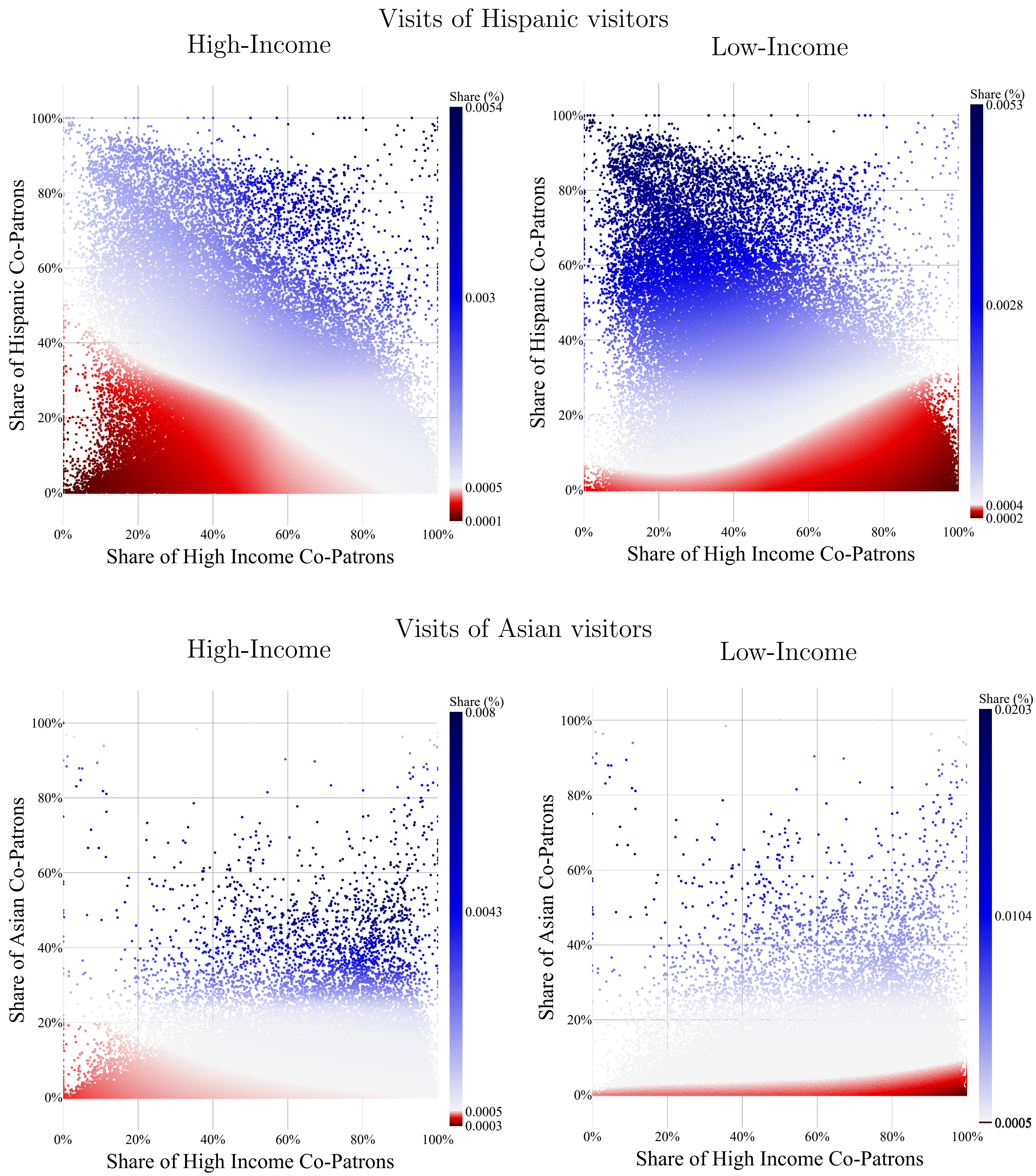


Table C.1: Mean exposure to own race co-patrons

	Low Income				High Income			
	White (1)	Black (2)	Hispanic (3)	Asian (4)	White (5)	Black (6)	Hispanic (7)	Asian (8)
All Categories	0.77	0.36	0.43	0.15	0.81	0.28	0.30	0.14
Restaurant	0.78	0.36	0.42	0.12	0.81	0.28	0.30	0.12
Bank	0.75	0.34	0.44	0.17	0.79	0.26	0.31	0.15
Bigbox	0.77	0.31	0.42	0.14	0.80	0.25	0.30	0.14
Convenience Store	0.78	0.45	0.41	0.12	0.81	0.35	0.30	0.12
Grocery Store	0.78	0.41	0.48	0.09	0.82	0.28	0.29	0.11
Gym	0.74	0.27	0.40	0.20	0.79	0.24	0.30	0.18
Pharmacy	0.76	0.41	0.45	0.21	0.81	0.30	0.31	0.15

NOTES: The table reports the average exposure to share of own-race co-patrons based on home-venue-home visits. Each row is computed by showing the average of share of own-race co-patrons for venues in the given chain category, weighted by visit share based on home-venue-home visits by each demographic group.

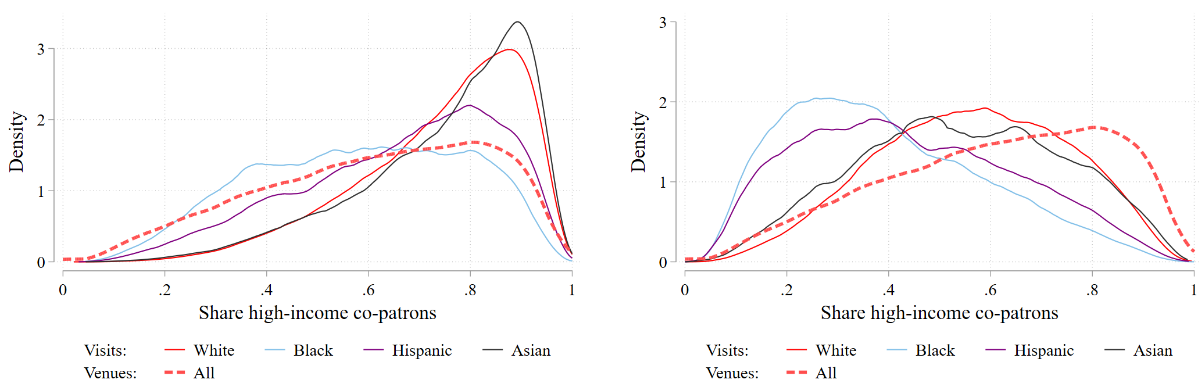
Table C.2: Mean exposure to high income co-patrons

	Low Income				High Income			
	White (1)	Black (2)	Hispanic (3)	Asian (4)	White (5)	Black (6)	Hispanic (7)	Asian (8)
All Categories	0.58	0.43	0.45	0.55	0.76	0.61	0.67	0.76
Restaurant	0.58	0.44	0.45	0.54	0.74	0.60	0.66	0.74
Bank	0.59	0.45	0.45	0.55	0.77	0.62	0.68	0.76
Bigbox	0.58	0.46	0.47	0.58	0.74	0.61	0.66	0.76
Convenience Store	0.52	0.35	0.41	0.49	0.71	0.54	0.63	0.71
Grocery Store	0.58	0.40	0.41	0.59	0.77	0.61	0.69	0.79
Gym	0.63	0.53	0.51	0.59	0.79	0.67	0.71	0.77
Pharmacy	0.57	0.40	0.41	0.51	0.77	0.60	0.68	0.76

NOTES: The table is analogous to Table C.1, but shows the average exposure to share of high-income co-patrons.

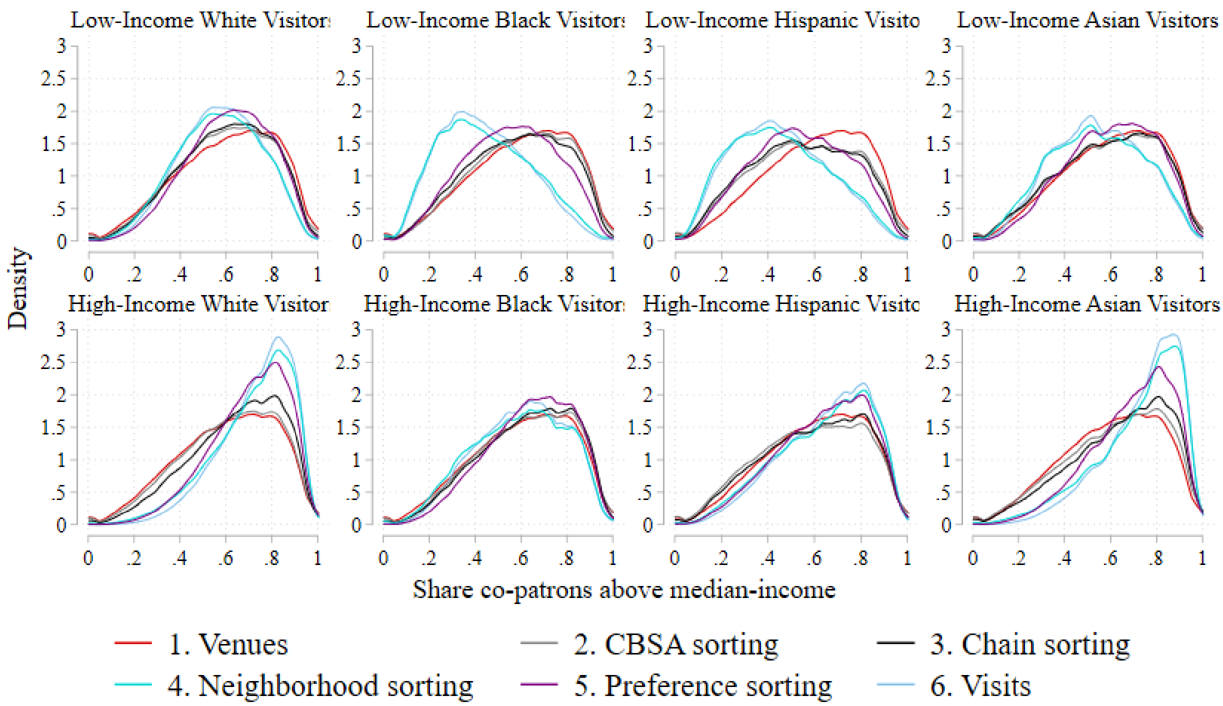
Figure C.2: Exposure to, and Availability of, High-Income Co-Patrons

All Chain Categories



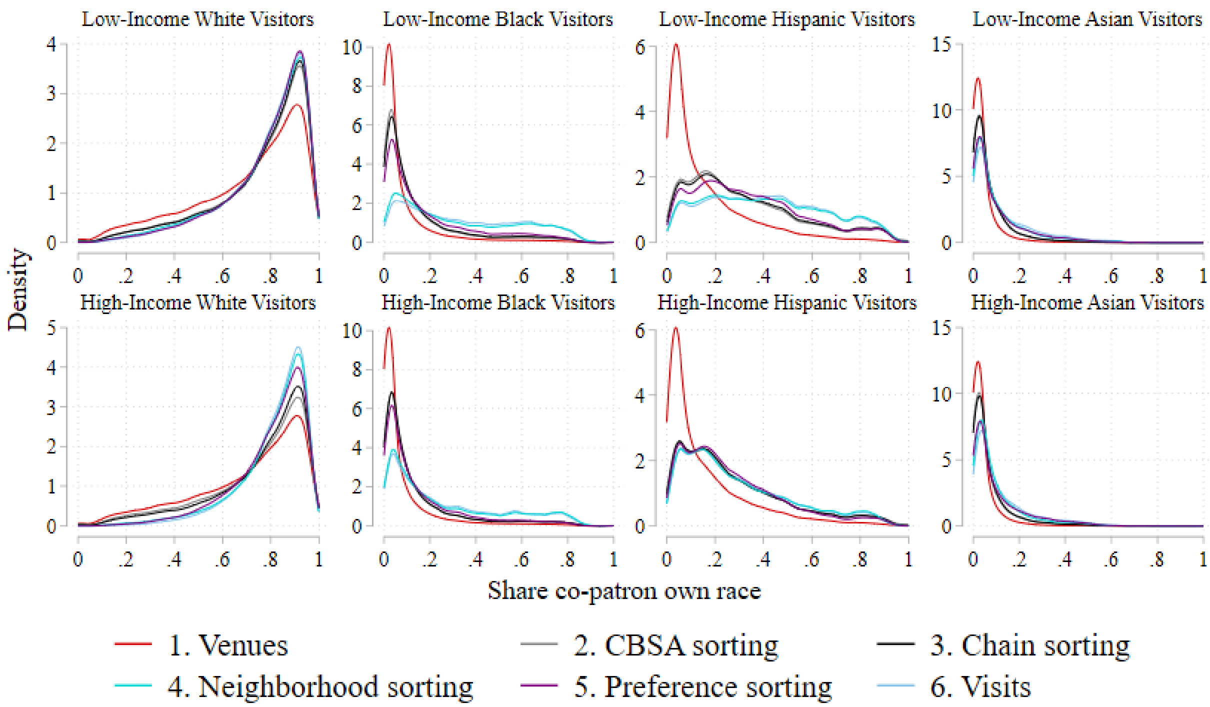
C.2 Determinants of social exposure

Figure C.3: Income decomposition



NOTES: The figure depicts kernel density plots of high-income co-patron shares of the venues visitors of a given income-race group visit under different scenarios. In the first scenario (1. Venues), the visitors visit all restaurant venues with equal probability. In the second (2. CBSA Sorting), visitors visit all restaurant venues in the CBSA where they reside with equal probability. In the third (3. Chain Sorting), visitors visit all restaurant venues in the CBSA where they reside and chains with probabilities proportional to their visits to each chain. In 4. Neighborhood Sorting, venues are weighted by counterfactual visit shares retaining between CBSA-chain nest variation and distance dis-utility, and in 5. Preference Sorting, venues are weighted by counterfactual visit shares retaining between CBSA-chain nest variation and preferences for co-patron characteristics. The final curve, 6. Visits, weights venues by the actual visit shares of a given group. Each plot reflects these kernel densities for a different income-race group.

Figure C.4: Race decomposition



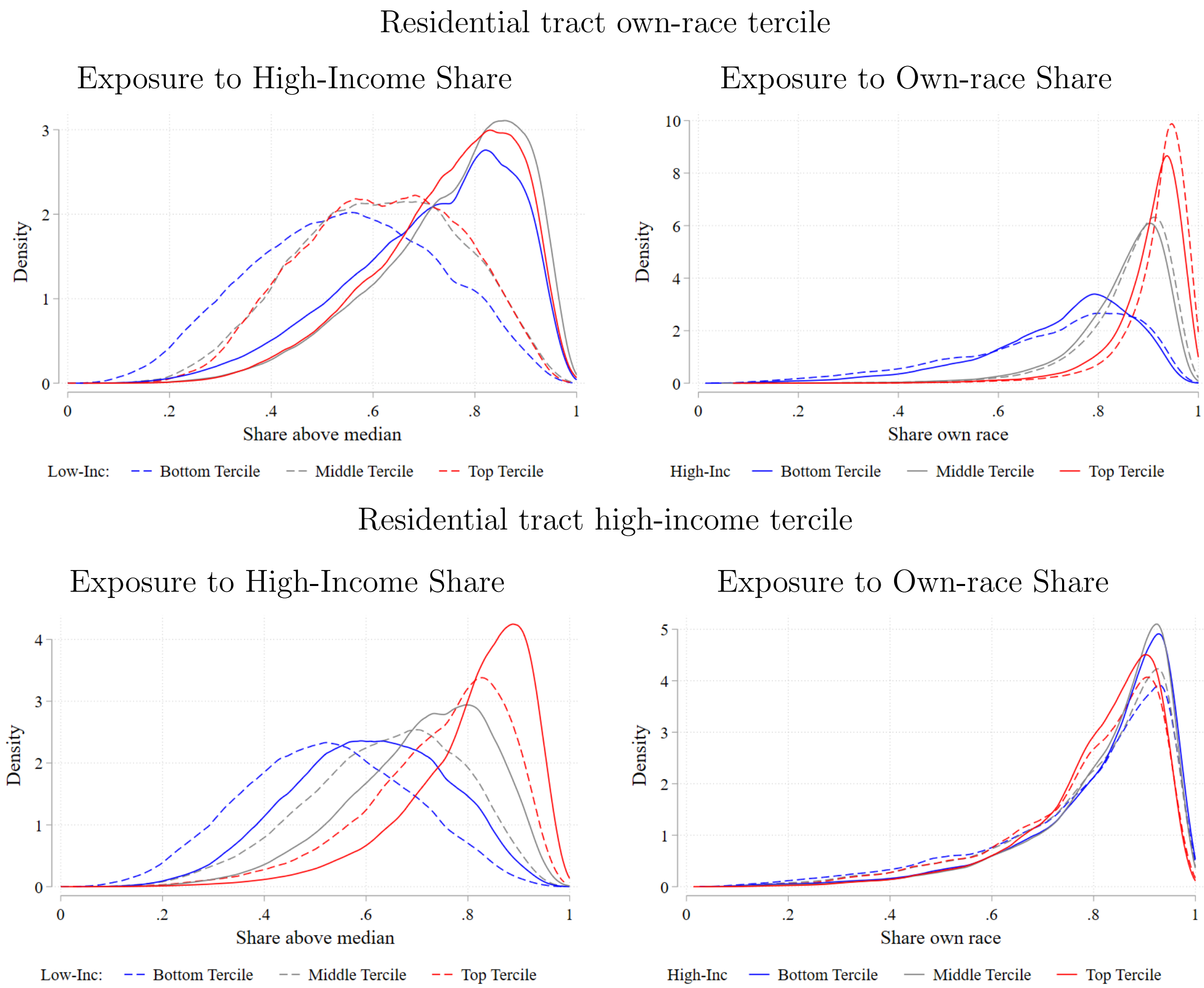
NOTES: The figure depicts kernel density plots of own-race co-patron shares of the venues visitors of a given income-race group visit under different scenarios. In the first scenario (1. Venues), the visitors visit all restaurant venues with equal probability. In the second (2. CBSA Sorting), visitors visit all restaurant venues in the CBSA where they reside with equal probability. In the third (3. Chain Sorting), visitors visit all restaurant venues in the CBSA where they reside and chains with probabilities proportional to their visits to each chain. In 4. Neighborhood Sorting, venues are weighted by counterfactual visit shares retaining between CBSA-chain nest variation and distance dis-utility, and in 5. Preference Sorting, venues are weighted by counterfactual visit shares retaining between CBSA-chain nest variation and preferences for co-patron characteristics. The final curve, 6. Visits, weights venues by the actual visit shares of a given group. Each plot reflects these kernel densities for a different income-race group.

Table C.3: Mean exposure to own race co-patrons

	Low Income				High Income			
	White (1)	Black (2)	Hispanic (3)	Asian (4)	White (5)	Black (6)	Hispanic (7)	Asian (8)
Venue	0.70	0.09	0.17	0.04	0.70	0.09	0.17	0.04
CBSA sorting	0.05	0.06	0.15	0.03	0.03	0.06	0.11	0.03
Chain sorting	0.05	0.08	0.16	0.03	0.05	0.06	0.11	0.03
Preference sorting	0.08	0.11	0.18	0.06	0.09	0.08	0.11	0.06
Neighborhood sorting	0.07	0.25	0.25	0.06	0.10	0.18	0.13	0.06
Income preference sorting only	0.03	0.07	0.14	0.03	0.05	0.04	0.08	0.03
Race preference sorting only	0.06	0.15	0.22	0.08	0.05	0.11	0.17	0.08
Model predicted visits	0.08	0.27	0.25	0.08	0.11	0.19	0.13	0.08
Actual visits (HVH)	0.08	0.27	0.25	0.08	0.11	0.19	0.13	0.08
Actual visits (All)	0.07	0.23	0.24	0.08	0.10	0.17	0.14	0.08

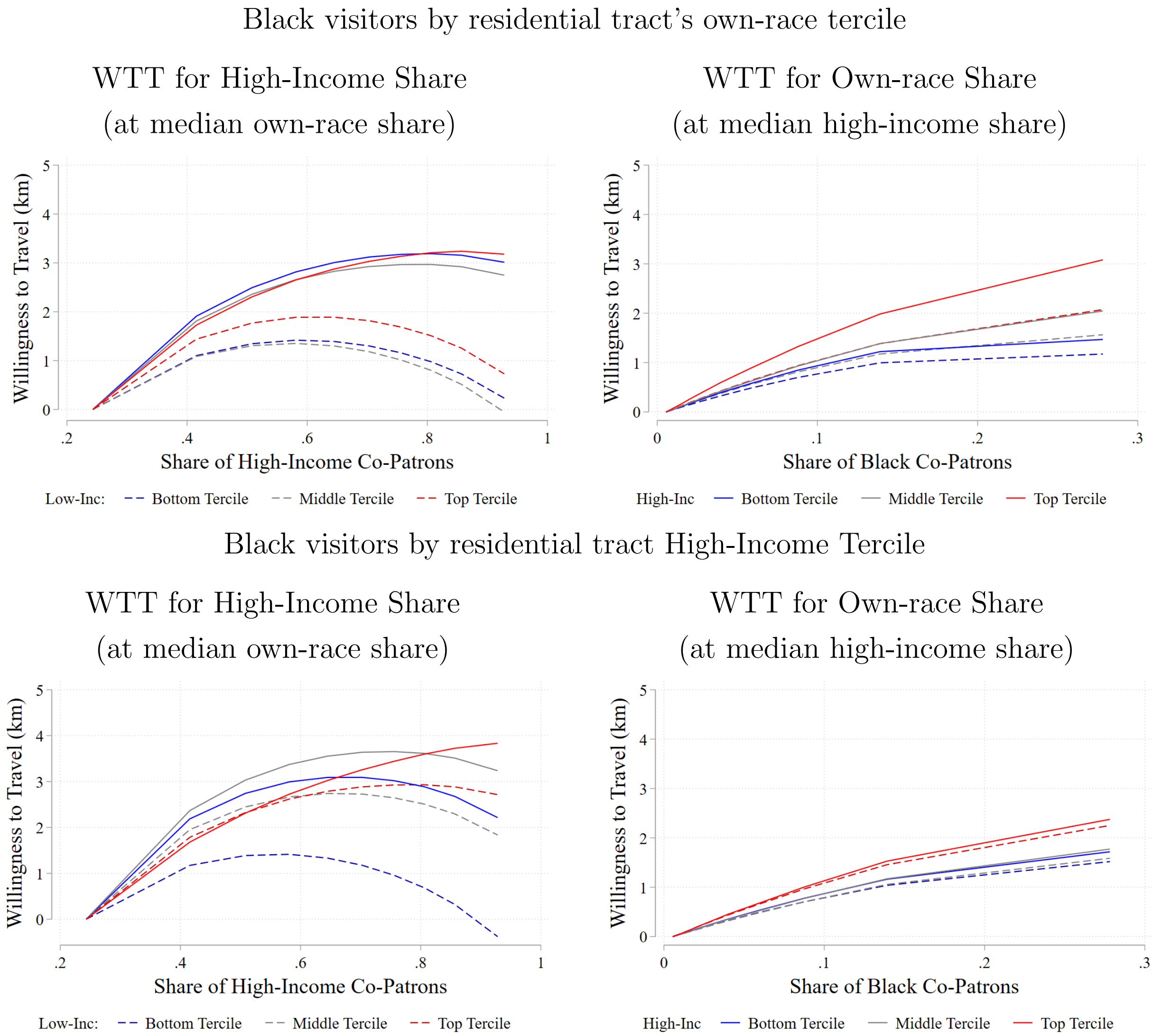
NOTES: This table is analogous to C.3 but reports mean exposure to own-race co-patrons. It shows the average share of own-race co-patrons an individual of each group would be exposed to under different counterfactual visit scenarios. All rows except the first row are adjusted relative to the "Venue" row. The first row evaluates the counterfactual scenario when devices visited venues uniformly at the national level. The second row evaluates the counterfactual scenario when devices visited venues uniformly within their CBSA of residence. The third row evaluates the counterfactual scenario when devices visited venues uniformly within their CBSA of residence and choice of chain. The fourth row evaluates the counterfactual scenario when devices consider their preferences for co-patron characteristics, in addition to their choice of CBSA and chain. The fifth row evaluates the counterfactual scenario when devices consider their distance dis-utility while ignoring preferences for co-patron characteristics, in addition to their choice of CBSA and chain. The sixth row evaluates the counterfactual scenario when devices only consider their preferences for own-race co-patrons, in addition to their choice of CBSA and chain. The seventh row evaluates the counterfactual scenario when devices only consider their preferences for high-income co-patrons, in addition to their choice of CBSA and chain. The eighth row evaluates the counterfactual scenario when devices consider both their preferences for co-patron compositions and distance dis-utility, in addition to their choice of CBSA and chain. The ninth row shows the actual exposure based on home-venue-home visits. Finally, the tenth row shows the actual exposure based on all types of visits.

Figure C.5: Spatial variation in exposure for White visitors



NOTES: This figure depicts white devices' exposure to share of high-income co-patrons and share of white co-patrons from home-venue-home visits by spatial heterogeneous groups. We split high-income white and low-income white devices into three groups based on the share of white residents in census tracts (the top panel) or the share of high-income residents in census tracts (the bottom panel). The terciles are created by splitting the share of white residents or share of high-income residents into terciles, weighted by the total population of census tracts.

Figure C.6: Spatial Heterogeneity

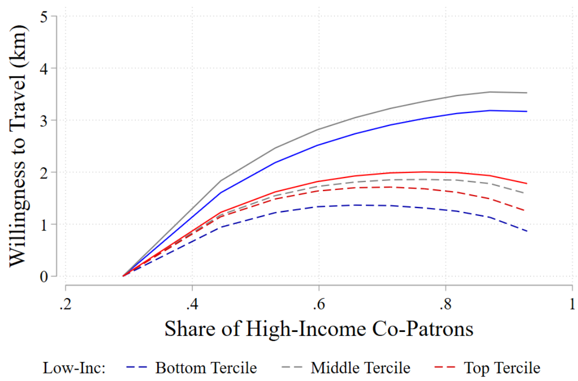


NOTES: Continued on next page. This figure is the heterogeneity version of Figure 4. This figure depicts the preference estimates of spatial heterogeneous groups by showing each dimension of co-patron composition while fixing the other at the median. We split each of the 8 demographic groups into three groups based on the share of white residents in census tracts (the top panel) or the share of high-income residents in census tracts (the bottom panel). The terciles are created by splitting the share of white residents or share of high-income residents into terciles, weighted by the total population of census tracts.

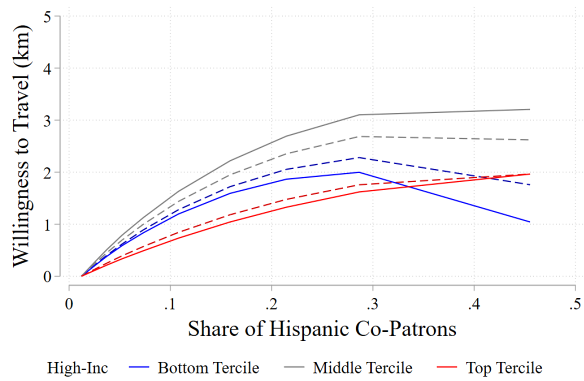
Spatial Heterogeneity. cont

Hispanic visitors by residential tract's own-race tercile

WTT for High-Income Share
(at median own-race share)

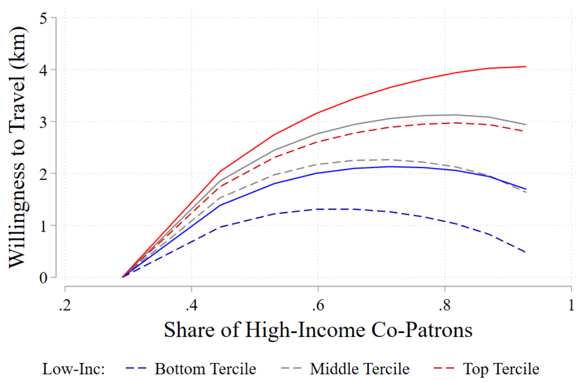


WTT for Own-race Share
(at median high-income share)

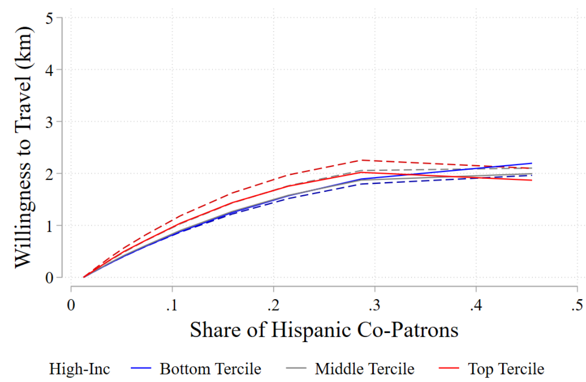


Hispanic visitors by residential tract's high-income tercile

WTT for High-Income Share
(at median own-race share)



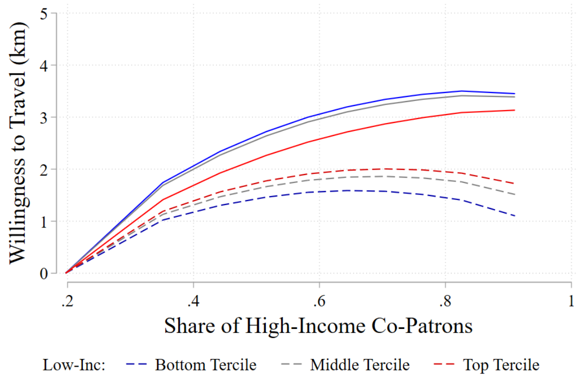
WTT for Own-race Share
(at median high-income share)



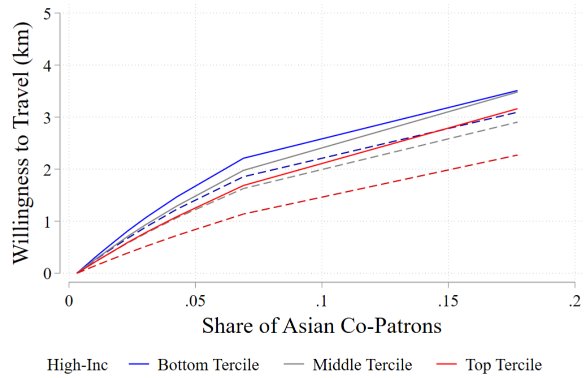
Spatial Heterogeneity. cont

Asian Visitors by residential tract's own-race tercile

WTT for High-Income Share
(at median own-race share)

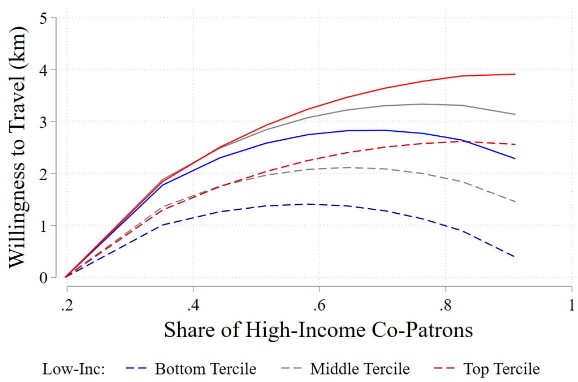


WTT for Own-race Share
(at median high-income share)



Asian Visitors by residential tract's high-income tercile

WTT for High-Income Share
(at median own-race share)



WTT for Own-race Share
(at median high-income share)

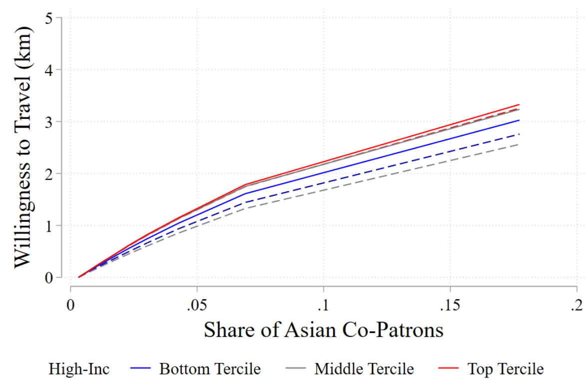
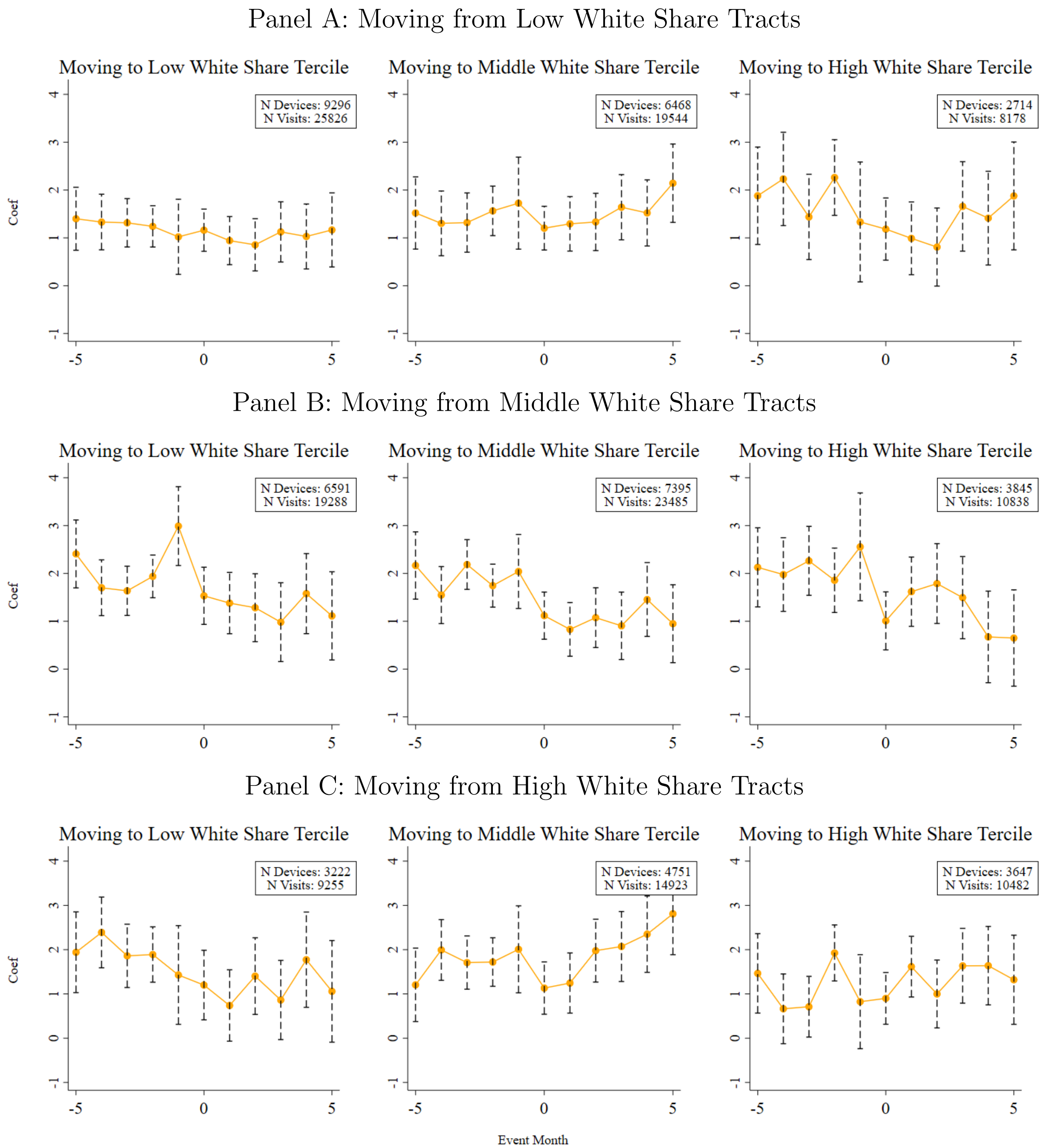


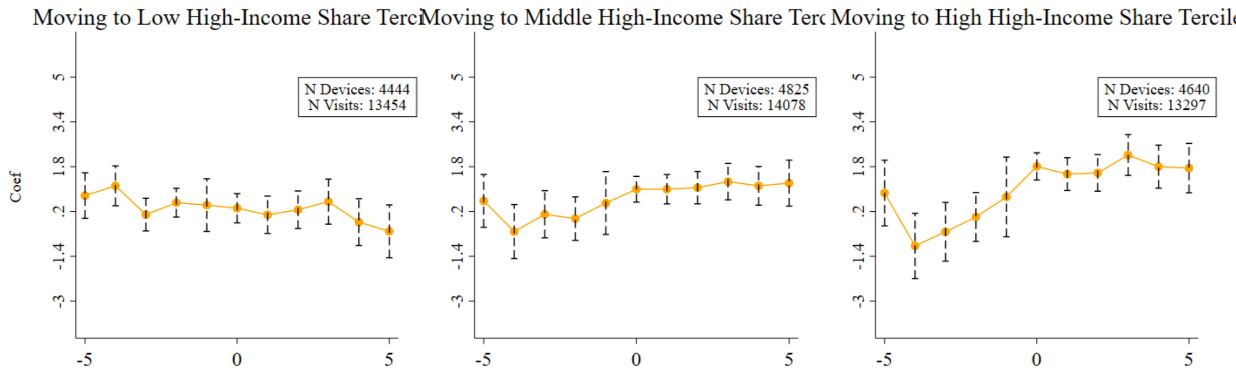
Figure C.7: Mover's Result Across Own Race Terciles: High income preference



Notes: Each plot in this figure presents the coefficients estimated in MLE with time-variant preferences over co-patrons and the cost of distance, and time-invariant chain dummies. Each point represents the coefficient on the own race co-patron share for a given month since moves occurred, with the bands reflecting 95% confidence intervals on those estimates. We sample home-venue-home visits to restaurants by only cross-CBSA high income white movers and split movers by origin-destination own race tercile pairs. We do not draw random samples of visit cases and retain all visits cases.

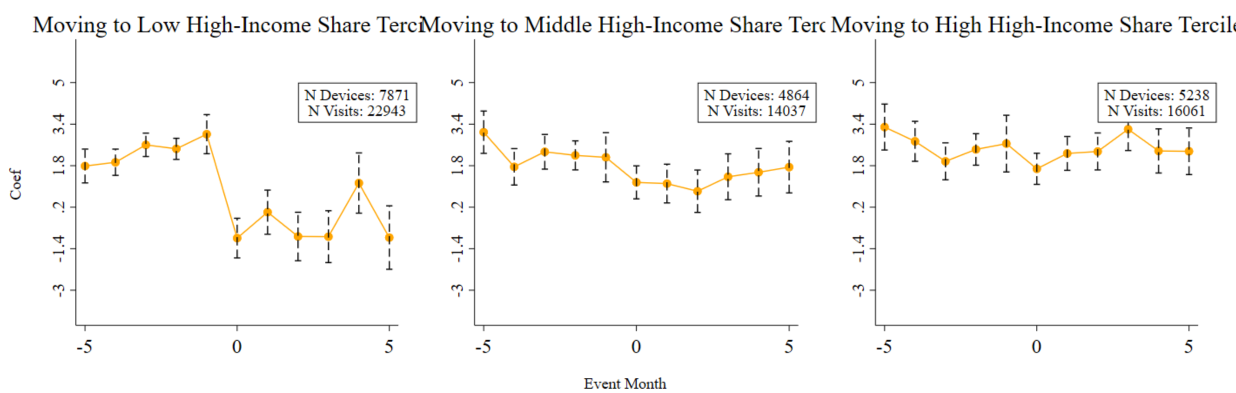
Figure C.8: Movers Result Across High-Income Terciles: High income preference

Panel A: Moving from Low High-Income Share Tracts



Panel B: Moving from Middle High-Income Share Tracts

Panel C: Moving from High High-Income Share Tracts

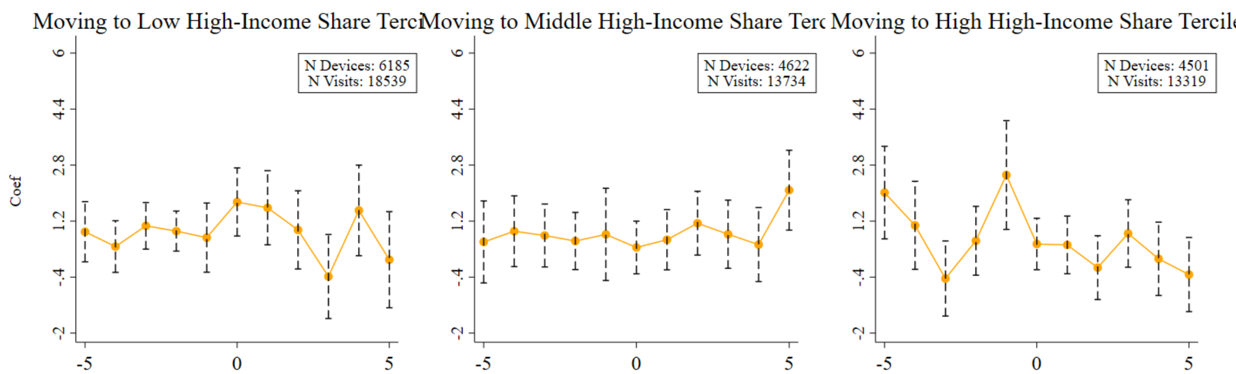


NOTES: Each plot in this figure presents the coefficients estimated in MLE with time-variant preferences over co-patrons and the cost of distance, and time-invariant chain dummies. Each point represents the coefficient on the own race co-patron share for a given month since moves occurred, with the bands reflecting 95% confidence intervals on those estimates. We sample home-venue-home visits to restaurants by only cross-CBSA high income white movers and split movers by origin-destination high-income tercile pairs. We do not draw random samples of visit cases and retain all visits cases.

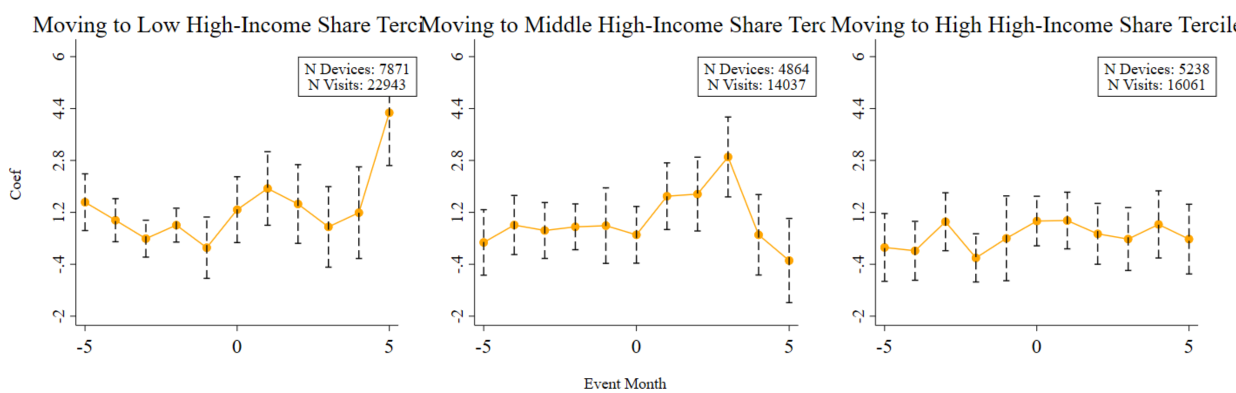
Figure C.9: Mover's Result Across High-Income Terciles: Own race preference

Panel A: Moving from Low High-Income Share Tracts

Panel B: Moving from Middle High-Income Share Tracts



Panel C: Moving from High High-Income Share Tracts



NOTES: Each plot in this figure presents the coefficients estimated in MLE with time-variant preferences over co-patrons and the cost of distance, and time-invariant chain dummies. Each point represents the coefficient on the own race co-patron share for a given month since moves occurred, with the bands reflecting 95% confidence intervals on those estimates. We sample home-venue-home visits to restaurants by only cross-CBSA high income white movers and split movers by origin-destination high-income tercile pairs. We do not draw random samples of visit cases and retain all visits cases.

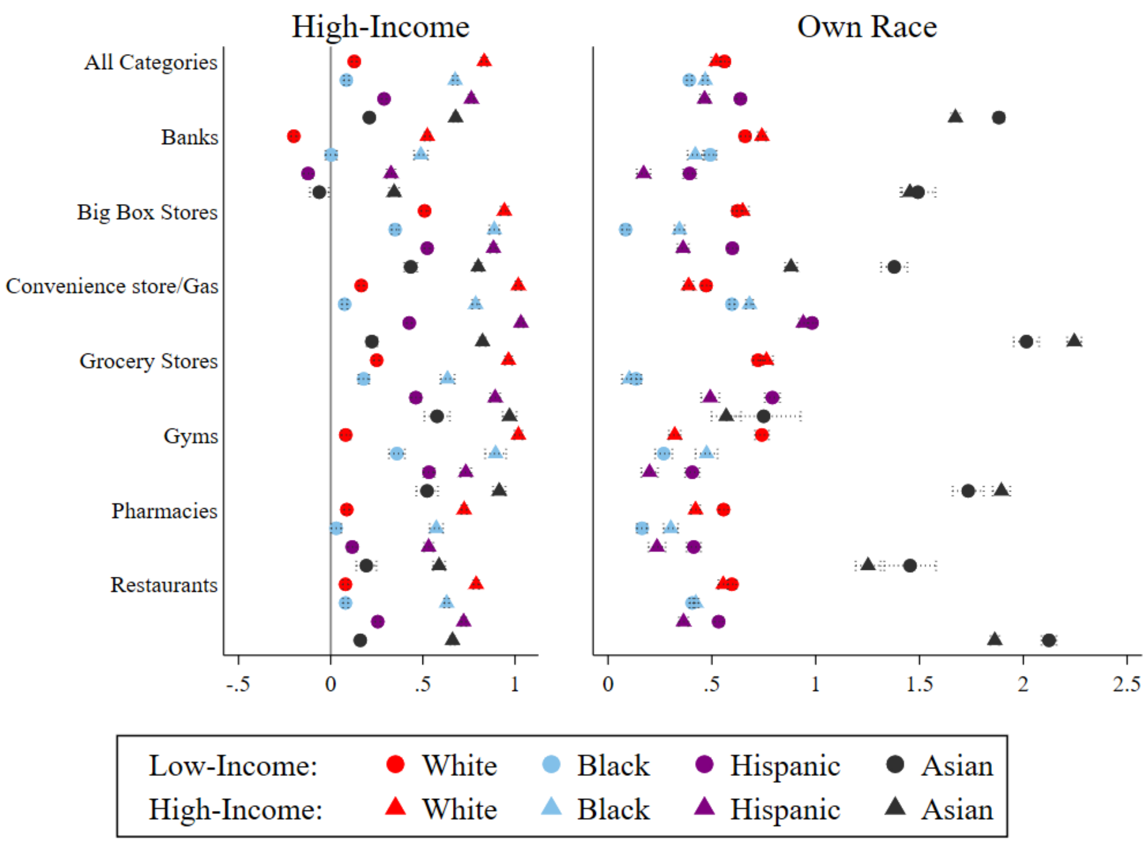
Table C.4: Mean coefficients before and after move

	O1			O2			O3		
	D1	D2	D3	D1	D2	D3	D1	D2	D3
Panel A: own race									
Pre-Move	1.73	1.88	1.07	0.82	0.75	1.16	0.76	0.65	0.25
	(1.11)	(1.40)	(1.61)	(0.87)	(1.18)	(1.40)	(0.82)	(1.05)	(1.12)
Post-Move	1.68	0.91	0.31	0.92	0.95	0.26	1.82	1.18	0.67
	(1.21)	(1.07)	(1.08)	(1.44)	(1.17)	(1.14)	(1.58)	(1.40)	(1.16)
Panel B: high income									
Pre-Move	0.58	0.12	0.01	1.42	1.23	1.10	2.36	2.30	2.61
	(0.82)	(1.05)	(1.29)	(0.62)	(0.87)	(1.06)	(0.64)	(0.84)	(0.93)
Post-Move	0.09	1.11	1.78	0.43	1.03	2.01	-0.45	1.29	2.37
	(0.92)	(0.77)	(0.85)	(1.11)	(0.85)	(0.91)	(1.23)	(1.03)	(0.93)

NOTES: This table shows the estimated preference coefficient on share of white co-patrons within venues, for high-income white individuals who moved permanently to a new CBSA. The coefficients are estimated in MLE with time-variant preferences over co-patrons and the cost of distance, and time-invariant chain dummies. Pre-move preferences average estimates over 5 months prior to the move. Post-move preferences average estimates over five months after the move. Standard errors are pooled across 5 months prior or post moving as well. We sample home-venue-home visits to restaurants by only cross-CBSA high income white movers and split movers by origin-destination high-income tercile pairs. We do not draw random samples of visit cases and retain all visits cases.

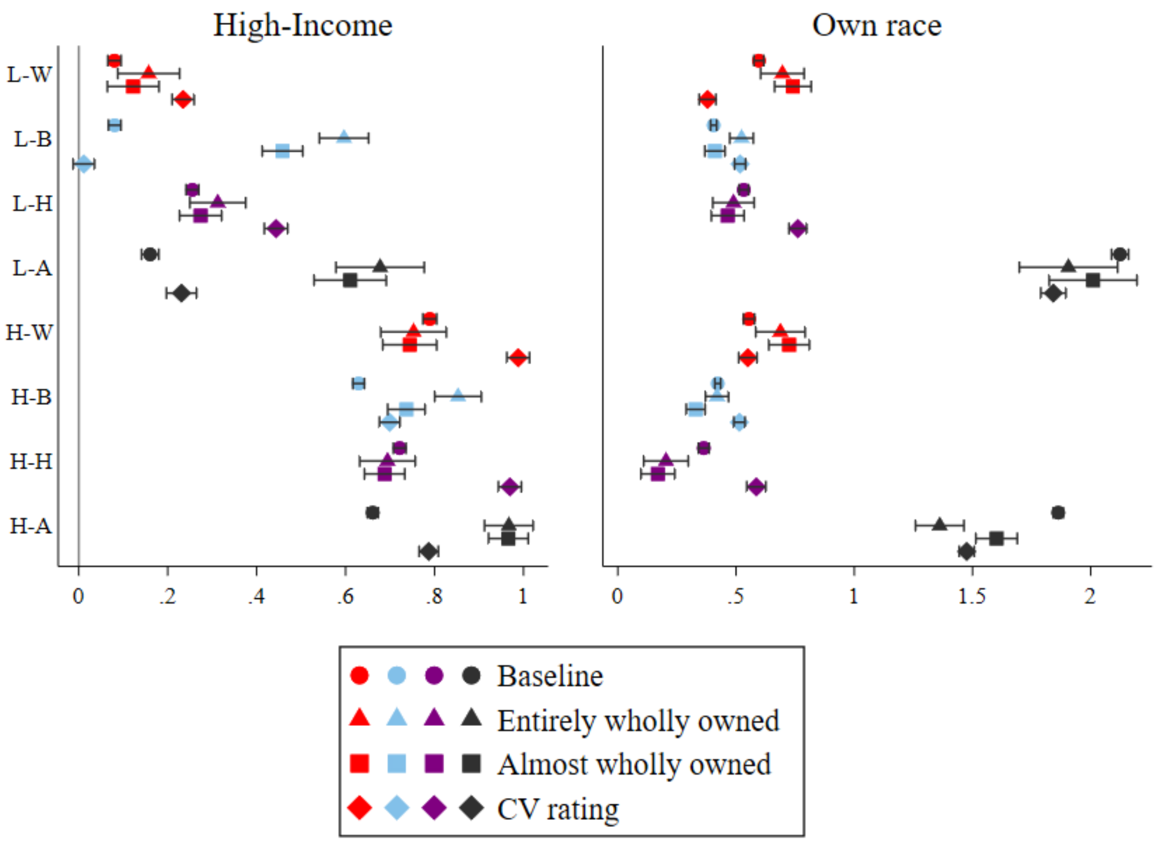
C.3 Robustness checks

Figure C.10: Robustness to different categories



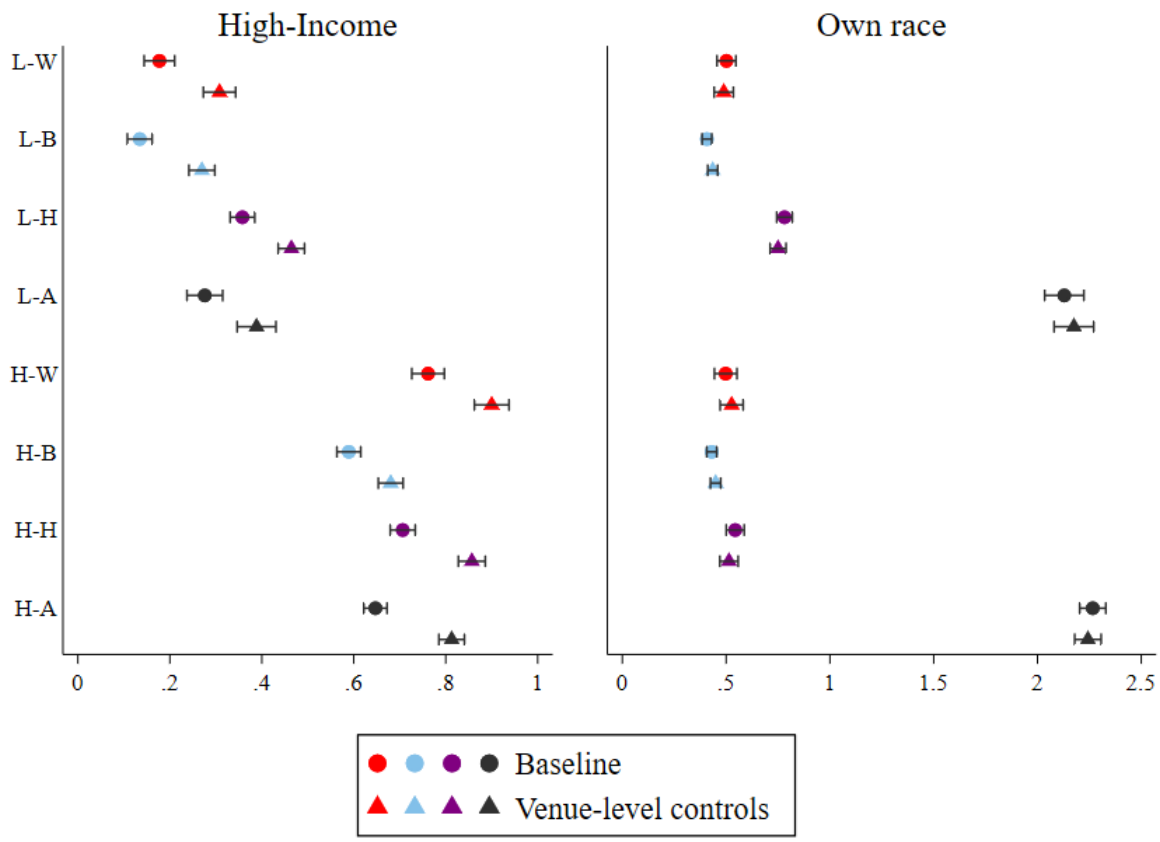
NOTES: The figures present the coefficients estimated in MLE with linear share of own race and share of high income co-patrons, linear and quadratic costs of distance. We draw random samples of 750,000 visit cases, if the total number of visit cases for the scenario exceeds 750,000. Each point represents the coefficient on the share of high-income co-patrons (left) or share of own race co-patrons (right) for a specific chain category, with the bands reflecting 95% confidence intervals on those estimates. Both the coefficients and standard errors are adjusted by distance elasticity at the the average distance of the chosen venues in the estimation sample.

Figure C.11: Robustness to standardized chains categories



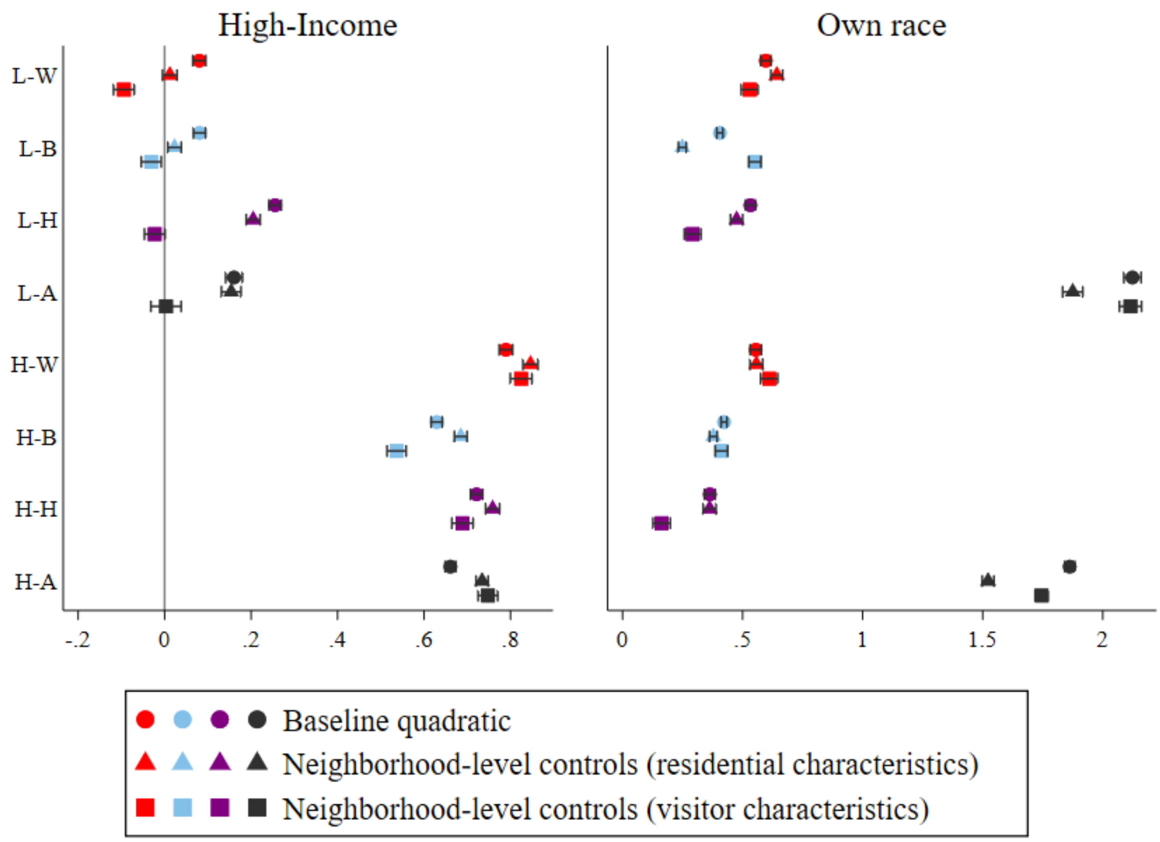
Notes: The figures are analogous to Figure C.10. Points marked by circle are estimated in MLE with linear share of own race and share of high income co-patrons, linear and quadratic costs of distance on random samples of 750,000 visit cases to restaurant chains. The baseline specification is the same as the restaurant specification in Figure C.10. Points marked by triangle are estimated on the baseline estimation sample while keeping only venues in entirely wholly owned chains, defined as 5% or fewer are franchised venues. Points marked by square are estimated on the baseline estimation sample while keeping only venues in almost wholly owned chains, defined as 20% or fewer are franchised venues. Points marked by diamond are estimated on the baseline estimation sample while keeping only top quartile venues with the lowest coefficient of variation in Google Places star rating. All coefficients and standard errors are adjusted by distance elasticity at the the average distance of the chosen venues in the estimation sample.

Figure C.12: Robustness to adding venue-level controls



NOTES: The figures are analogous to C.10. Points marked by circle are estimated in MLE with linear share of own race and share of high income co-patrons, linear and quadratic costs of distance on random samples of 750,000 visit cases to restaurant chains. The baseline is the same as the restaurant specification in C.10. Points marked by triangle are estimated in MLE with linear share of own race and share of high income co-patrons, linear and quadratic costs of distance, Google Place star rating, Google Place number of review, and venue square footage. The two specifications estimate on the same estimation samples from the baseline specification. All coefficients and standard errors are adjusted by distance elasticity at the the average distance of the chosen venues in the estimation sample.

Figure C.13: Robustness to adding neighborhood-level controls



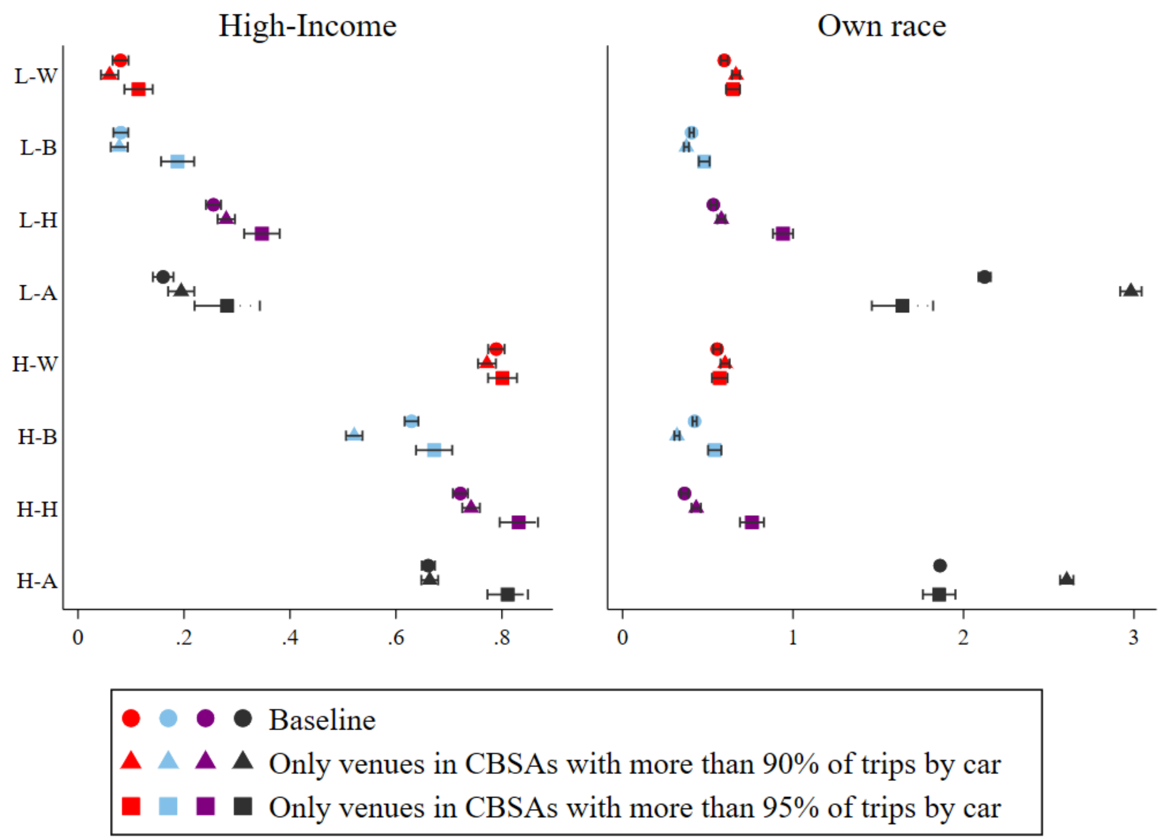
Notes: The figures are analogous to C.10. Points marked by circle are estimated in MLE with linear share of own race and share of high income co-patrons, linear and quadratic costs of distance on random samples of 750,000 visit cases to restaurant chains. The baseline is the same as the restaurant specification in C.10. Points marked by triangle are estimated in MLE with linear share of own race and share of high income co-patrons, linear and quadratic costs of distance, share of own race residents and share of high income residents in the census tract where each venue is located. Points marked by square are estimated in MLE with linear share of own race and share of high income co-patrons, linear and quadratic costs of distance, share of own race co-patrons and share of high income co-patrons to all other commercial venues within the census tract where each venue is located. The three specifications estimate on the same estimation samples from the baseline specification. All coefficients and standard errors are adjusted by distance elasticity at the the average distance of the chosen venues in the estimation sample.

Table C.5: Coefficients on visitor shares, residential share controls, and visitor to other venue controls

		High Income				Own Race			
Specification:		**Resident Spec**		**Visitor Spec**		**Resident Spec**		**Visitor Spec**	
Covariates:		Co-patron Share (1)	Resident Share (2)	Co-patron Share (3)	Other Venue Share (4)	Co-patron Share (5)	Resident Share (6)	Co-patron Share (7)	Other Venue Share (8)
Low-Income	White	.023	.364	-.183	.393	1.245	-.101	1.027	.136
		(.016)	(.018)	(.023)	(.025)	(.024)	(.013)	(.034)	(.038)
Low-Income	Black	.043	.319	-.058	.243	.466	.313	1.035	-.3
		(.015)	(.016)	(.022)	(.023)	(.015)	(.01)	(.024)	(.025)
Low-Income	Hispanic	.385	.266	-.043	.714	.893	.089	.547	.764
		(.015)	(.017)	(.022)	(.024)	(.024)	(.015)	(.032)	(.037)
Low-Income	Asian	.284	-.054	.006	.318	3.452	.82	3.908	.123
		(.021)	(.027)	(.032)	(.033)	(.038)	(.03)	(.042)	(.054)
High-Income	White	1.716	-.265	1.674	-.113	1.13	.004	1.24	-.122
		(.018)	(.019)	(.025)	(.028)	(.027)	(.014)	(.036)	(.041)
High-Income	Black	1.398	-.229	1.096	.22	.771	.107	.841	.044
		(.015)	(.016)	(.022)	(.024)	(.016)	(.011)	(.026)	(.028)
High-Income	Hispanic	1.522	-.163	1.382	.122	.728	.011	.326	.602
		(.016)	(.017)	(.025)	(.027)	(.027)	(.017)	(.036)	(.042)
High-Income	Asian	1.437	-.585	1.468	-.228	2.98	1.006	3.425	.65
		(.014)	(.017)	(.022)	(.023)	(.024)	(.017)	(.026)	(.035)

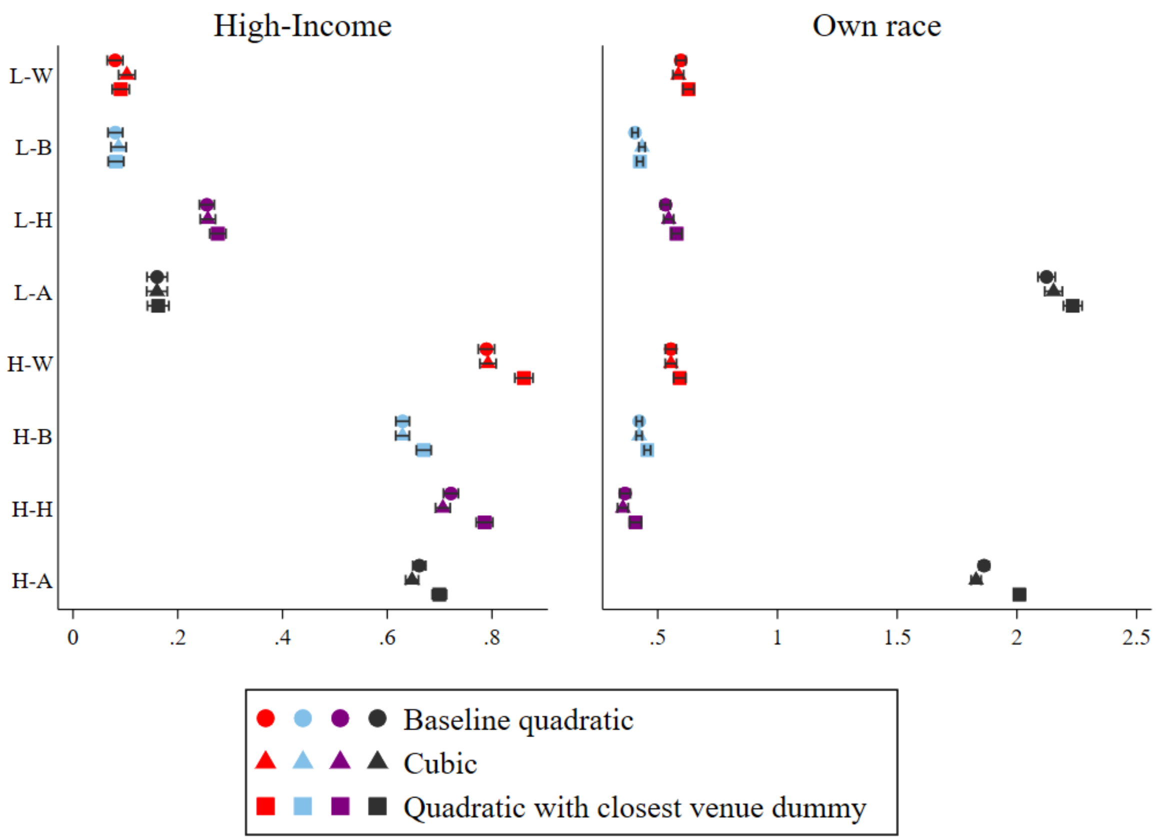
NOTES: This table presents the preference coefficients of co-patron compositions while controlling for neighborhood characteristics, prior to distance elasticity adjustments. The resident specification shows the same specification as points marked by triangle in Figure C.13. The visitor specification shows the same specification as points marked by square in Figure C.13.

Figure C.14: Robustness to transportation mode



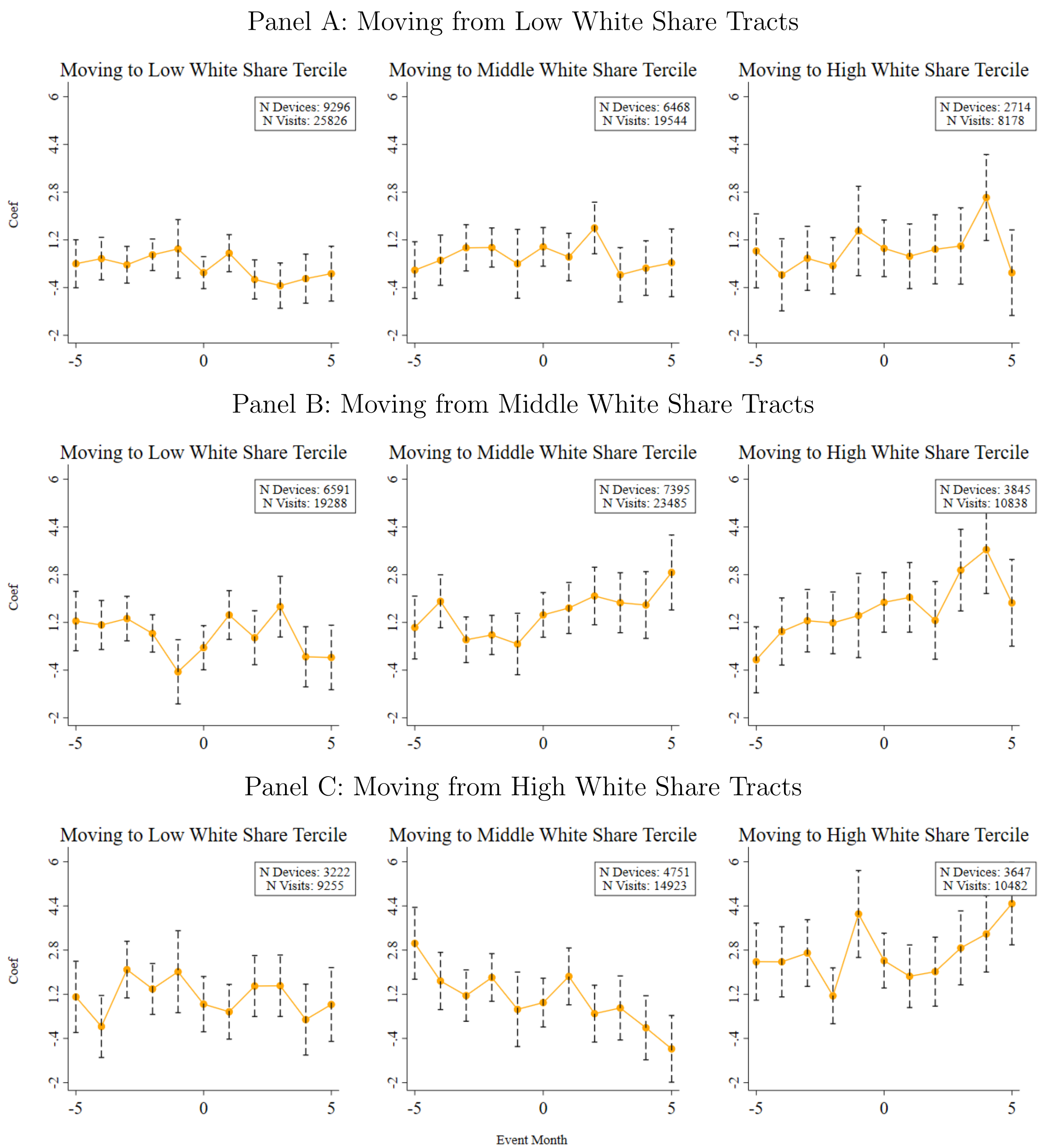
NOTES: The figures are analogous to C.10. Points marked by circle are estimated in MLE with linear share of own race and share of high income co-patrons, linear and quadratic costs of distance on random samples of 750,000 visit cases to restaurant chains. The baseline specification is the same as the restaurant specification in C.10. Points marked by triangle are estimated on the baseline estimation sample while keeping only venues in CBSAs with more than 90% of trips to commercial consumption venues by car. Such selection procedure results in , Points marked by square are estimated on the baseline estimation sample while keeping only venues in CBSAs with more than 95% of trips to commercial consumption venues by car. All coefficients and standard errors are adjusted by distance elasticity at the the average distance of the chosen venues in the estimation sample. +
For each of the 8 demographic groups, the estimation samples for the baseline specification include devices residing in top 100 CBSAs. If we only keep CBSAs with more than 90% of trips by car, then 79 CBSAs remain. If we only keep CBSAs with more than 95% of trips by car, then 23 CBSAs remain.

Figure C.15: Robustness to distance specifications



Notes: The figures are analogous to C.10. Points marked by circle are estimated in MLE with linear share of own race and share of high income co-patrons, linear and quadratic costs of distance on random samples of 750,000 visit cases to restaurant chains. The baseline is the same as the restaurant specification in C.10. Points marked by triangle are estimated in MLE with linear share of own race and share of high income co-patrons, linear, quadratic and cubic costs of distance. Points marked by square are estimated in MLE with linear share of own race and share of high income co-patrons, linear and quadratic cubic costs of distance, and a dummy indicating whether the venue is the closest to the device within chain. The four specifications estimate on the same estimation samples from the baseline specification. All coefficients and standard errors are adjusted by distance elasticity at the the average distance of the chosen venues in the estimation sample.

Figure C.16: Own Race Preference of High Income White Devices Before & After Moves



Notes: Each plot in this figure presents the coefficients estimated via MLE with time-varying preferences over co-patrons and the cost of distance, and time-invariant chain dummies. Each point represents the coefficient on the own race co-patron share for a given month since moves occurred, with the bands reflecting 95% confidence intervals on those estimates. We sample home-venue-home visits to restaurants by only cross-CBSA high income white movers and split movers by origin-destination own race tercile pairs. We do not draw random samples of visit cases and retain all visits cases.