Demographic Preferences and Income Segregation*

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

January 2025

Abstract

We study how preferences over the demographic composition of co-patrons affects income segregation in shared spaces. To distinguish demographic preferences from tastes for other venue attributes, we study venue choices within business chains. We find two notable regularities: preferences for high-income co-patrons are similar across racial groups, and racial homophily does not vary by income. These demographic preferences are economically large, explain much of the cross-group variation in exposure to high-income co-patrons, and correlate with movers' neighborhood choices.

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

*We are grateful to Fahad Algifari, Andrea Cerrato, Hongrui He, Shih-Hsuan Hsu, Luigi Roncoroni, John Ruf, and Serena Xu for excellent research assistance. We thank Drew Breunig, Nicholas Shailas, Stephanie De Leon, Elizabeth Cutrone, and the team at Precisely PlaceIQ for data access and helpful conversations. We thank Cody Cook for sharing his block-level income estimates. We thank Keith Chen, Pierre Magontier, and numerous seminar participants for helpful feedback. The views expressed herein are those of the authors and do not necessarily reflect the views of Precisely 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 and the University of Pennsylvania under protocol 856399. 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 Toronto
‡Columbia University, NBER, and CEPR
§Princeton University
¶University of Pennsylvania and NBER

1 Introduction

American life is segregated by income and race in domains ranging from residences to media diets. Given that social connections matter for economic mobility, the segregation of interactions by income raises important questions. Whether people's preferences over the demographic composition of those around them contribute to income segregation is particularly contentious. In this paper, we study how different demographic groups are exposed to high-income individuals in shared commercial spaces. We estimate individuals' preferences over the racial and income composition of co-patrons and use them to quantify sources of cross-group differences in experienced income segregation.

To measure exposure to high-income co-patrons, we use data on the movements of millions of smartphones in the United States in 2018 and 2019. Joining these movement data with building-level residential demographics, 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 (Asian, Black, Hispanic, and White) interacted with two income categories (split by median income).[1]

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 Asian and White individuals. High-income Black individuals, for example, experience nearly the same high-income exposure as low-income White individuals.

We consider three explanations for these demographic differences in experienced income segregation. First, differences in proximity to venues: low-income individuals may live far from venues with high-income patrons. Second, differences in preferences for product attributes: groups might vary in their price sensitivity or taste for particular services. Third, preferences over the demographics of co-patrons: these preferences encompass all the ways co-patron mix may affect an individual's likelihood of choosing a venue. This includes their affinities for certain groups but also, for instance, how a concentration of young professionals working in a coffee shop might create a productive ambiance.

To distinguish between these explanations, we estimate preferences using choices of venues within chain businesses. In our venue-choice model, people trade off the cost of a longer trip with the benefits of venue characteristics. Controlling for proximity, we separate preferences over co-patron mix from tastes for product attributes by contrasting venues within the same

[1] Following US government and social science conventions, the four racial/ethnic groups we observe are non-Hispanic Asian, non-Hispanic Black, Hispanic, and non-Hispanic White. For the sake of brevity, we often omit the adjective "non-Hispanic" and simply refer to these racial-ethnic groups as racial groups.

chain, which offer the same product but vary in their co-patron demographics.[2] We specify demographic preferences as a flexible function of the high-income share of co-patrons and the same-race share of co-patrons. This flexibility allows us to capture complementarities between race and income composition and to distinguish preferring perfectly homogeneous co-patrons from a mere aversion to being unlike everyone else.

Our baseline estimation sample contains visits to restaurants, the business category with the largest number of chain venues. Chain restaurants are well suited as a laboratory to estimate demographic preferences, because they are frequently visited and tend to have homogeneous venues. We can therefore estimate demographic preferences, and their evolution over time, with greater precision and detail than one could in other choice settings like home purchases or school enrollments. To the extent possible, we also report demographic exposure and preference estimates within a range of other commercial and public venues.

Our estimates reveal notable regularities across demographic groups in their preferences over co-patron composition. High- and low-income individuals exhibit similar levels of racial homophily (a preference for one's own race). Black, Hispanic, and White individuals have similar levels of racial homophily (with that of Asian individuals being somewhat stronger). Members of different racial groups have broadly similar preferences for high-income exposure (with those of White individuals being somewhat weaker). Only high-income individuals, however, exhibit monotone preferences over the share of co-patrons who are high-income. Low-income individuals prefer establishments with an integrated mix of low- and high-income co-patrons.

These preferences for demographic exposure are economically large. Individuals are willing to travel two to three additional kilometers to visit a venue in the 95$^{\text{th}}$ percentile of either the same-race or high-income distribution rather than a venue at the 5$^{\text{th}}$ percentile.[3] 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; Bayer, Ferreira, and McMillan, 2007).

Given these regularities in demographic preferences, why does demographic exposure differ between groups? Differences in tastes for product attributes, the first of three possible explanations, are small: little income segregation arises from high-income individuals visit-

[2]For retail chains, brand power and economies of scale depend on a standardized offering, so product availability and service quality in a chain venue do not typically reflect the local composition of co-patrons. Even chains whose design footprint is less standardized, like Starbucks, strive to create a consistent experience and fixed menu across venues. In a robustness exercise, we restrict our estimation sample to the most standardized restaurant chains, such as Olive Garden, which is wholly owned instead of franchised and whose Google review ratings vary little across venues.

[3]The exception to this 95$^{\text{th}}$-5$^{\text{th}}$ comparison is low-income individuals' preference for high-income exposure, because they prefer economically integrated venues.

ing different chains. In fact, within-chain income segregation resembles income segregation across all non-residential venues, including public and commercial venues. For example, both high- and low-income people visit McDonald's restaurants, but they tend to choose different McDonald's locations.[4] We use our estimated model to show that neighborhood of residence and demographic preferences are the factors that largely explain cross-group differences in exposure.

Racial differences in high-income exposure stem from how preferences over co-patron demographics interact with the joint distribution of income and race across venues. For instance, high-income Black and White individuals have 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 distances between their residences and high-income venues. This reflects the role of racial homophily. Because majority-Black and majority-Hispanic venues generally have lower-income co-patrons, Black and Hispanic individuals face a trade-off between visiting heavily high-income venues and visiting heavily same-race venues that Asian and White individuals do not face.

The gap in high-income exposure between low- and high-income individuals within racial groups reflects differences in residential sorting and preferences. Low-income people both live in poorer neighborhoods and have weaker preferences for high-income co-patrons. High-income individuals tend to live in neighborhoods near venues with many high-income co-patrons. Conditional on where they live, their stronger income preferences lead them to choose venues with more high-income co-patrons. Overall, demographic preferences explain observed income exposure more for high- than low-income people.

To further examine how demographic preferences relate to neighborhood choice, we estimate how demographic preferences vary across neighborhoods with different demographic mixes. We find that people live in neighborhoods that match their preferences for demographic exposure: within demographic groups, 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' demographic preferences and the dominant demographics of their residential neighborhoods suggests that demographic 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 by examining movers. Specifically, we estimate the preferences of individuals before and after they move between neighborhoods in different metropolitan

[4]This aligns with Cook (2023), who notes the widespread popularity of large chains, with McDonald's being the most preferred restaurant across all income groups.

areas. The estimated preferences are consistent with people sorting into neighborhoods based on their preferences over co-patron demographics. For instance, those moving to integrated neighborhoods show lower racial homophily before the move, even when controlling for their origin neighborhood's demographics. Consistent with intergroup contact theory, preferences for the local demographic mix strengthen after the move.[5] This mover analysis also validates our model specification: the estimated preferences do not shift discontinuously when an individual's choice set changes. Although this investigation of movers' preferences is limited by smaller sample sizes and a short time horizon, it demonstrates the potential for mobility data to advance our understanding of demographic preferences.

The main contribution of this paper is to a debate over the existence and importance of demographic preferences. These preferences are believed to play an important role in explaining residential segregation (e.g., Schelling 1971; Card, Mas, and Rothstein 2008), but a key challenge is to distinguish preferences over neighbors' income and racial demographics from tastes for other neighborhood amenities (Caetano and Maheshri, 2021; Bayer et al., 2022; Davis, Gregory, and Hartley, 2023; Schönholzer, 2023; Li, 2023).[6] By studying venue choice instead of neighborhood choice, we offer the first estimate of demographic preferences in shared spaces. The business chains we study also have considerably more uniform attributes than residential neighborhoods, and we observe many more choices.[7]

We also contribute to a literature documenting segregation in non-residential domains. Economists have documented the 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; Magontier, Schläpfer, and von Ehrlich, 2022; Cook, 2023; Massenkoff and Wilmers, 2023; Nilforoshan et al., 2023; Yabe et al., 2023), or student status (Cook, Currier, and Glaeser, 2022). We document experienced segregation within commercial spaces by income and race jointly, using building-level demographic information.[8]

[5] The intergroup contact hypothesis was originally formulated by Allport (1954). Recent work testing this hypothesis includes Lowe (2021), Cantoni and Pons (2022), and Bursztyn et al. (2024).

[6] In addition to demographic preferences and differences in tastes for neighborhood amenities, persistent racial segregation of residences has also been attributed to wealth differences, prejudice, and housing-market discrimination (Charles, 2003; Rothstein, 2017).

[7] In related work, Davis et al. (2019) show that Yelp users in New York City are more likely to visit restaurants in neighborhoods with racial composition similar to their own. Backstrom and Woodward (2021) show that anglers are less likely to fish from a site with larger Black and Hispanic populations than their neighborhood.

[8] In this vein, Wang et al. (2018) show that residents of Black and Hispanic neighborhoods visit high-income neighborhoods less despite traveling as far as others. In the housing market, racial differences in

Finally, our paper quantifies the drivers of income segregation in shared spaces, complementing 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). Chetty et al. (2022a,b) find that one's number of high-socioeconomic-status Facebook friends is among the strongest predictors of economic mobility. Our smartphone movement data, however, measure social exposure, not social connections. Atkin, Chen, and Popov (2022) use similar 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. Choi, Guzman, and Small (2024) show that the introduction of Starbucks into US neighborhoods with no coffee shops increases entrepreneurship. 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 income segregation and estimate demographic preferences, we need to know the demographic characteristics and consumption trips of a large sample of individuals. This section describes the construction of our estimation sample from smartphone movement data and building-level demographic data. Appendix A offers more details on each data source.

2.1 Data sources

Our smartphone movement data are from Precisely PlaceIQ, a location data and analytics firm. Precisely PlaceIQ aggregates pings from applications that request locational services from the smartphone's operating system.[9] 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 to buildings, 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. Precisely PlaceIQ uses the timing of the first and last ping in the visit ping cluster to compute a lower bound for visit duration.

income mean that racial minorities face a trade-off between sorting into high same-race and high-income neighborhoods (Sethi and Somanathan, 2004; Bayer, Fang, and McMillan, 2014; Bruch, 2014; Reardon, Fox, and Townsend, 2015). We show that minorities face the same trade-off when choosing daily social spaces. Like some studies of residential decisions (Bayer and McMillan, 2005; Aliprantis, Carroll, and Young, 2022), we find that racial homophily plays an important role in this setting.

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

The demographic characteristics of each device come from building-level data that include the income bracket and race of individuals living at an address. Precisely PlaceIQ does not disclose the third-party provider of these data, so we discuss the reliability of this 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 assign demographics to devices based on their inferred residence, which is the residential building where the device regularly spends time at night (Couture et al., 2021).

2.2 Estimation sample

This subsection describes the selection of devices, venues, and visits in our analysis. To estimate preferences, we create a restricted sample of devices and visits for which we confidently know demographic information and trip purpose. To measure 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 have exactly one inferred residence over our 19-month sample period.[10] 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 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 Asian, Black, Hispanic, and White. We use visits by all devices for which we have building-level data to measure the demographic mix of copatrons in each venue, applying device demographics probabilistically. To estimate group-specific preferences, we limit our estimation sample to the 36 million devices that reside in buildings in which at least two-thirds of the residents belong to the same income and racial group. 91% of buildings are racially homogeneous and 99% of buildings are economically homogeneous, consistent with most Americans living in single-family homes and significant sorting by residents of multi-family dwellings.

Venue selection criteria Our baseline estimation sample uses trips to restaurants, which have by far the largest number of chains, establishments, and visits. We also characterize co-patron exposure and estimate demographic preferences in banks, big box retail stores, convenience stores & gas stations, grocery stores, gyms, and pharmacies.

[10] Around 10% of devices move during this period. We drop these from our baseline estimation sample. We return to studying these movers in Section 6.2.2.

Table A.4 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, with a low of 70% 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 measurement concerns related to venue 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 the estimation sample.

Visit selection criteria Our estimation sample is restricted to trips to a venue that originate and end at home.[11] Considering only these direct trips ensures that their sole purpose is visiting a venue. This selection eliminates confounding factors due to trip chaining and allows us to identify preferences within a venue-choice model like that we introduce in Section 4. It means the estimation sample only includes devices that ping frequently enough to track direct trips. We also exclude visits with duration longer than three hours, as these are likely by venue employees.[12] Overall, the sample of restaurant visits we use to estimate 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. To estimate the demographic composition of these venues, we use the sample of all 1.5 billion trips to these restaurants. While we rely on a narrower sample of visits to estimate preferences, we later show that visit patterns in our estimation sample mirror those from much broader samples.

2.3 Data quality and representativeness

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

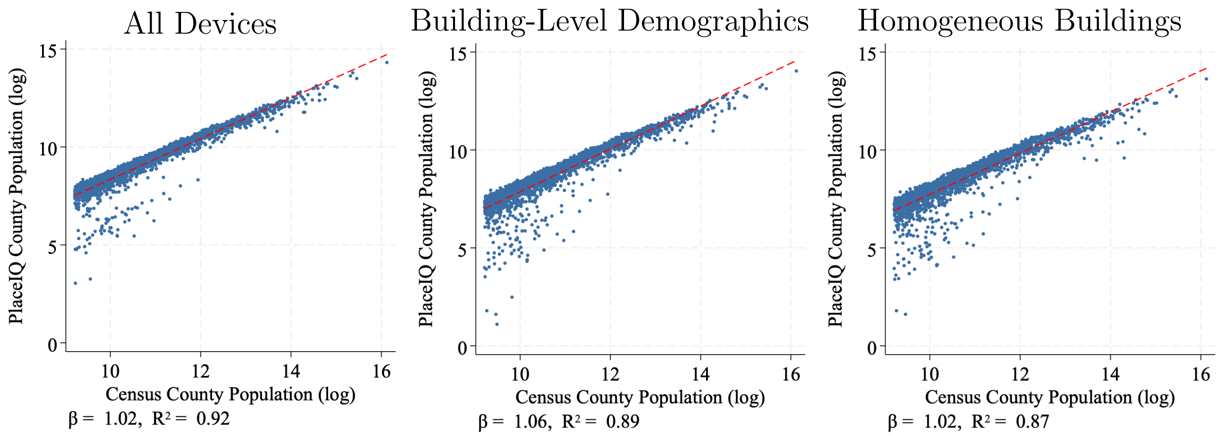
Couture et al. (2021) show that devices active in the smartphone data are broadly spatially representative and make visits that resemble what travelers self-report in the National Household Travel Survey. Figure 1 shows that the additional selection criteria we impose on our estimation sample involve only limited spatial biases. Panel A plots the number of

[11] We define a direct trip as a visit to a venue where the preceding and succeeding visits were to a device's home and within the same "activity day" (a 24-hour period starting at 3:00 AM). Since not every stop at a venue is observed, some trips may be mis-categorized as direct. Davis et al. (2019) and Miyauchi, Nakajima, and Redding (2021) study consumption trips that can originate at workplaces.

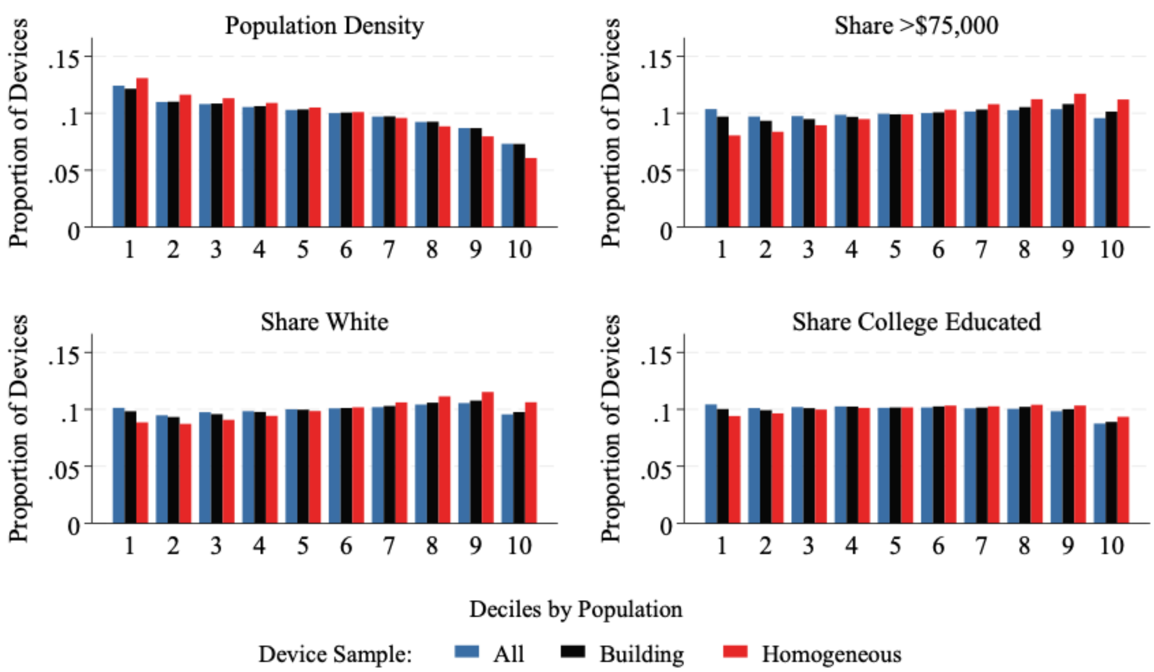
[12] Visit duration is measured with error, but can be reasonably interpreted as a lower bound for actual duration. A visit is registered when a smartphone application collects a ping in a venue, not when the smartphone first enters the venue, so a device may spend more time at a venue than we observe.

Figure 1: Comparing estimation sample to full smartphone data

Panel A: County population



Panel B: Population by within-county deciles



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 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 ($> \$75,000$) residents, population share of White residents, and population share of residents who have obtained a bachelor's degree. Panel B reports these decile shares for the three populations of devices shown in Panel A.

devices residing in a county for three device samples against the 2019 Census county population estimates. The "All Devices" sample includes all devices that have exactly one home assignment. The "Building" sample includes only devices with building-level demographic data. The "Homogeneous Building" sample includes only devices living in buildings in which one income-race group constitutes more than 67% of residents. Regressing the device count of a county on its Census population count yields an R^2 of at least 0.87 for each of the three device samples. Our estimation sample is nearly as representative of county population as all smartphone data.

Panel B of Figure 1 evaluates the spatial representativeness of our sample within counties. For instance, we show that within a county, we have about the same number of devices in block groups with the highest White share as in block groups with the lowest White share. Specifically, we compute the share of devices living in each decile of population density, high-income share, White share, and college-educated share, defined within each county using block-group data from the 2015-2019 American Community Survey. If device samples were exactly proportionate to populations in the American Community Survey within each county, each bar would be of equal height (0.10). We show these results for the three different device 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 attention to homogeneous buildings, as we do in the estimation sample of devices, we see a more substantial bias away from the densest 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. Thus, 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.

Figure 1 demonstrates that our smartphone device samples are spatially representative of residences using Census characteristics, but Appendix A.4 shows that the building-level data still contain more White and high-income households. That said, the cross-county correlation between the share of a given demographic group in the Census and the share of devices in that group based on our building-level data remains above 0.8 for all demographic groups (Figure A.1). 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) by reporting differences in demographic exposure across groups instead of absolute levels that may overstate exposure to high-income devices.

Finally, Appendix A.5 validates our building-level demographic information in three ways. In each exercise, we show that the building-level data predict individual characteristics or behavior *within* the smallest geography for which Census data are available. First, we compare the incomes in our building-level data to those of Cook (2023), which imputes income using home parcel characteristics from CoreLogic. These two sources agree whether a Census block has above-median income for 75% of blocks. The discrepancies are modest: treating blocks within $10,000 of the median as matching raises the agreement to 91%. These block-level measures are significantly correlated within block groups, suggesting that our building-level data reveal income variation below the smallest geographic level for which the Census publishes income statistics.

Second, we use voter registration data from North Carolina to show that the building-level demographic information predicts the race of voters at a given address better than one could using Census data. Our building-level data matches the race of Black voters 20% more often than one could by randomly drawing households in the same Census block group - the smallest geography for which both race and income are available - and 5% more often than one could using Census block data on race.

Third, we run an internal validation check: we show that the building-level demographic information predicts differences in the venue choices of residents of different buildings in the same Census block group. For example, residents of high-income buildings are more likely than their low-income neighbors to visit chains preferred by high-income people, such as Starbucks.[13] This final exercise delivers two conclusions that we leverage in our empirical analysis. First, behavior predicted using the building-level demographic data is consistent with behavior predicted using Census demographic data. Second, the building-level data provides more information than the Census, because it allows us to predict variation in behavior within the smallest geographic unit for which both income and race are available in the Census.

3 Demographic exposure in shared spaces

This section documents how exposure to different types of co-patrons varies by demographic group. We first report, for each demographic group, the full distribution of visits to chain

[13]To make such comparisons, we first calculate each restaurant chain's relative popularity among high-income versus low-income patrons within the same tract, using block-group-level income assignments. We then compute each chain's relative popularity at the building level, contrasting high- and low-income buildings in the same block group. These block group-level and building-level measures of a chain's relative popularity with high-income patrons have a rank correlation of 0.8 (Figure A.2). We find similarly high correlations for racial instead of income groups and for convenience stores & gas stations (the second largest establishment category).

restaurant venues by racial and income co-patron mix. We then show how these differences in visit patterns translate into disparities in income exposure across groups. Finally, we show that other categories of venues and visits have similar levels of income segregation.

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 chain restaurant within the 100 largest metropolitan statistical areas (MSAs). We compute visit propensity using a non-parametric kernel regression of a demographic group's share of visits to a venue on its co-patron characteristics.

These plots show notable patterns in visit propensity across demographic groups. For all groups, visit propensity is increasing in same-race share. Within each race, higher-income individuals visit venues with greater shares of high-income co-patrons than their low-income peers. However, the distribution of available co-patron demographics varies starkly across racial groups. Unsurprisingly, many more venues are predominately White than Asian, Black, or Hispanic. Heavily White venues vary significantly in their income composition, whereas heavily Black and Hispanic venues tend to be low-income. There are very few venues with a high Asian share of co-patrons, regardless of co-patron income. These visit patterns echo familiar patterns of residential segregation by race and assortative matching by income (e.g., Bayer and McMillan 2005).

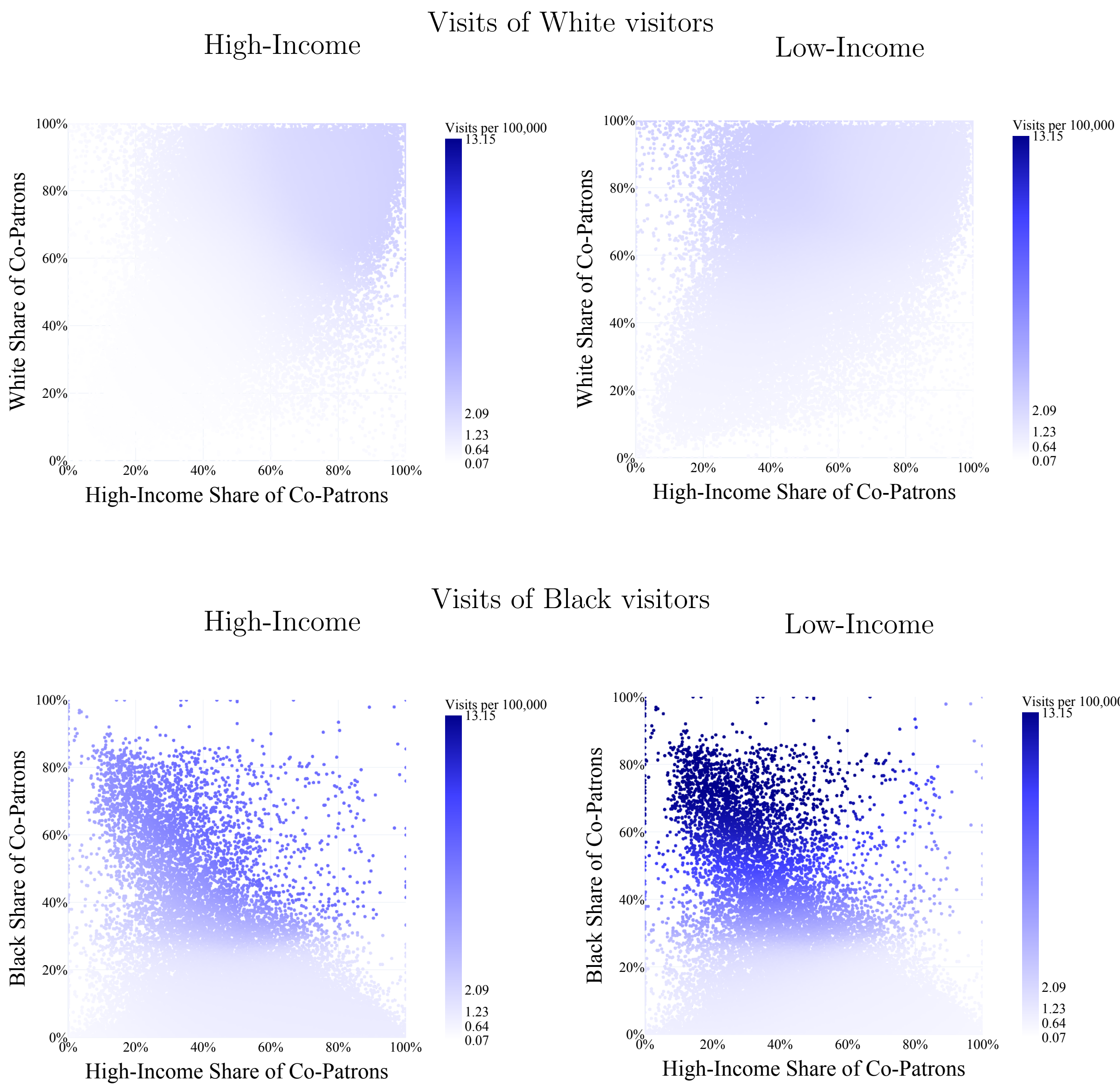
Table 1: Exposure to high-income co-patrons

	Low Income				High Income			
	Asian (1)	Black (2)	Hispanic (3)	White (4)	Asian (5)	Black (6)	Hispanic (7)	White (8)
Estimation sample	-0.06	-0.17	-0.15	-0.03	0.14	-0.00	0.05	0.13
All chain-restaurant visits	-0.04	-0.13	-0.13	-0.02	0.12	-0.00	0.04	0.12
All chain-venue visits	-0.05	-0.16	-0.14	-0.03	0.12	-0.01	0.04	0.11
All non-residence venue visits	-0.03	-0.15	-0.14	-0.03	0.18	0.04	0.09	0.17
All McDonald's restaurant visits	-0.08	-0.16	-0.16	-0.03	0.09	-0.02	0.02	0.11
Census tracts	0.04	-0.12	-0.08	-0.01	0.20	0.03	0.06	0.11

NOTES: This table reports, for different visit samples, the high-income share of co-patrons that each demographic groups (eight columns) is exposed to, relative to a baseline in which all venues in that sample are visited with uniform probability. The first row shows those high-income shares for visits in the estimation sample. The second through fourth rows shows those shares for broader visit samples. The fifth row shows those shares only for visits to McDonald's restaurants. In the sixth row, those shares are computed as if each Census tract is a venue and individuals only visit the census tract that they live in.

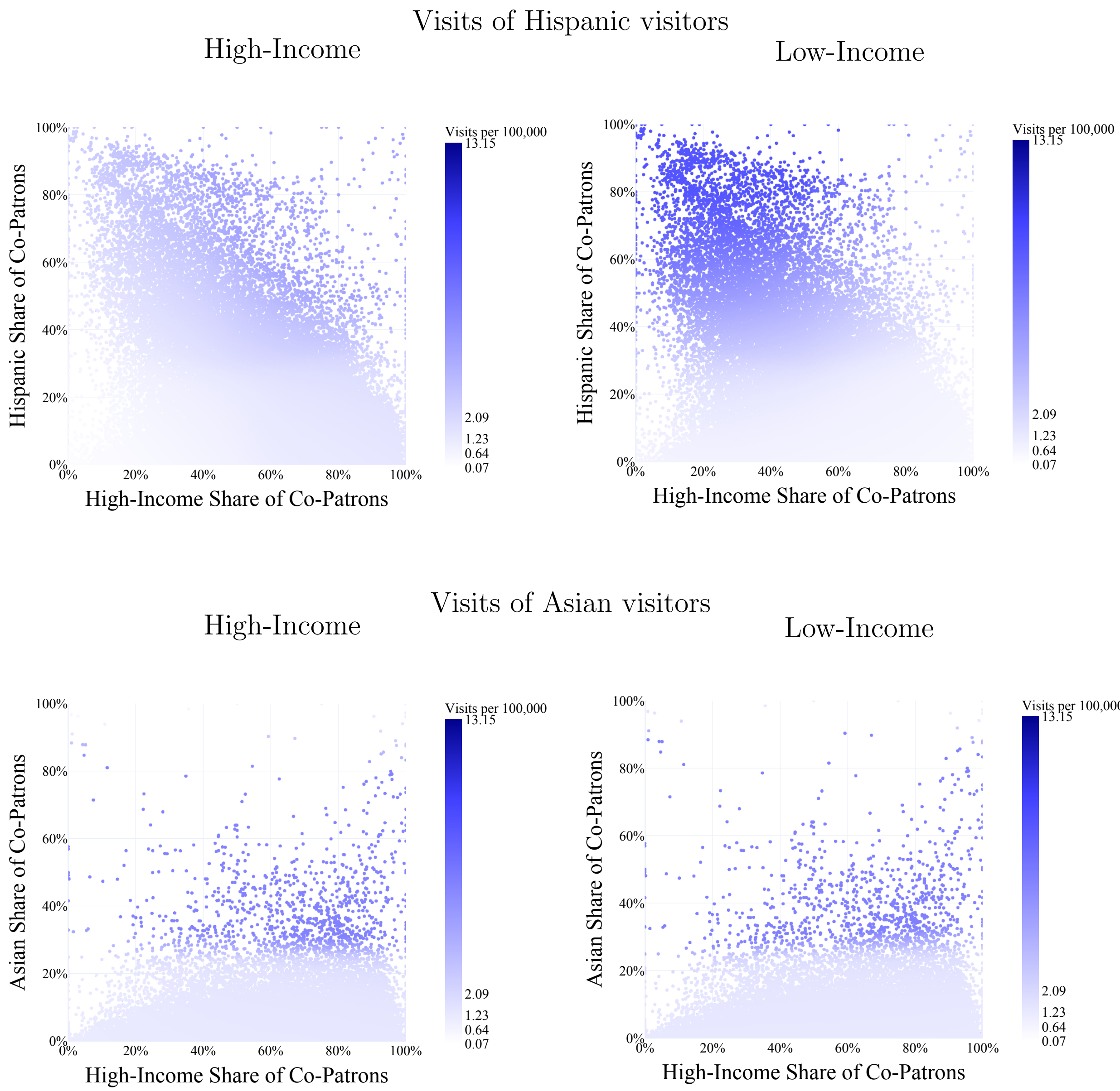
How do these differences in visit propensity and venue availability translate into differ-

Figure 2: Exposure to high-income and same-race co-patrons



Notes: This figure presents the results of a kernel regression of visit shares on co-patron race and income characteristics. Each panel displays the visit propensities for a specific demographic group, with each dot representing an individual restaurant. The color gradient is consistent across groups and follows a log-linear scale. This gradient is chosen to match the group with the greatest range in visit propensities between the most and least visited venues. We use an Epanechnikov kernel with a bandwidth of 0.05 and winsorize visit propensities at the 99th percentile for each group. *Continues on next page.*

Figure 2: Exposure to high-income and same-tace co-patrons (continued)



ences in income exposure? Table 1 reports exposure to high-income co-patrons for each demographic group across different environments. The first row reports high-income exposure computed using visits from our estimation sample. We chose this sample to suit our empirical strategy, not for its representativeness, so in subsequent rows we expand our sample to larger sets of visits and venues. The last row 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 income segregation in shared spaces compares with residential segregation.

Each column reports the income exposure of a given demographic group relative to a baseline in which these venues are visited uniformly. For example, column 8 of row 1 shows that high-income White individuals in our estimation sample choose restaurants with a share of high-income co-patron that is 13 percentage points higher than if they visited restaurants at random. We report our results in relative terms because absolute exposure levels are more sensitive to definitions of income group and sample biases.[14]

Table 1 yields two main results. First, there are substantial differences in exposure to high-income co-patrons across incomes and races. Within each race, high-income individuals have more high-income exposure: differences in mean exposure between high- and low-income individuals are typically 15 to 20 percentage points. Within each income group, Asian and White individuals have more high-income exposure than Black and Hispanic individuals. In fact, the average high-income exposure of a low-income White individual is only about 2 percentage points lower than that of a high-income Black individual. Second, we observe similar cross-group disparities in income exposure in a variety of settings. In particular, income segregation experienced within chain restaurants, based solely on visits in our estimation sample, resembles that experienced within all non-residential venues, both commercial and public. These broad income segregation patterns even manifest within a single chain like McDonald's. There is an approximately 25 percentage point difference between the high-income share in McDonald's locations visited by low-income Black or Hispanic individuals and those visited by high-income Asian or White individuals. These disparities in income exposure within shared spaces also mirror disparities in residential income exposure, computed using census tract demographic shares.[15]

[14]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 extensive margin, from variation in the number of visits that each demographic group makes. Smartphone samples with a partial history of each device's movements are not well-suited to study these differences. The National Household Transportation Survey, for instance, shows that despite differences in high- and low-income individuals' propensity to visit different destination types – high-income people make more trips to restaurant – high- and low-income people make similar numbers of consumption trips and similar numbers of non-work trips.

[15]Appendix Table B.1 shows similar regularities across environments for racial segregation.

A number of factors may explain this variation in exposure to high-income co-patrons across demographic groups. First, different groups may have different tastes 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 exposure to high-income co-patrons. Preferences for same-race co-patrons may generate differences in high-income exposure because venues with high-income co-patrons tend to be heavily White (Figure 2). In what follows, we investigate the relative importance of these explanations for demographic 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 over co-patron demographics and travel distances. We then describe how to use the estimated model to compute counterfactual venue visit shares that quantify the contributions of various mechanisms to cross-group variation in 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. We partition the set of venues into C disjoint subsets denoted by B_c (chains). Following Train (2009, Ch 4.2), we 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 multinomial logit case. The probability that decision maker i chooses venue 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, $\sum_i \sum_j I_{ij} \ln P_{j|ic}$, alone. 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.[16] We allow Y_{ij} to depend on observable attributes with coefficients that are common across chains. In order to identify parameters common across chains, we assume a common within-chain correlation of idiosyncratic preference shocks such that $\lambda_c = \lambda \; \forall c$.

[16]Chain-level attributes Y_{ic} that are common across establishments within a chain, such as menu items and prices, do not affect $P_{j|ic}$:

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

4.3 Mean utility specification

In the baseline specification, we assume that the mean utility of patronizing a venue within a chain depends on the distance to the venue and its co-patron composition. These preferences may vary across demographic groups, indexed by g. Preferences over distance 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 high-income share of co-patrons (s_j^{highinc}), and the same-race share of co-patrons (s_j^{samerace}):

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

where $f_1(\ln \text{distance}_{ij}; \delta^g)$ is a polynomial of log distance, $f_2(s_j^{\text{samerace}}, s_j^{\text{highinc}}; \beta^g)$ is a polynomial of the two co-patron shares, and δ^g and β^g are group-specific coefficient vectors on travel distance and co-patron composition, respectively.

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:

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

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. We define group g's willingness to travel for the co-patron composition $(s^{\text{samerace}}, s^{\text{highinc}})$ as the incremental distance Δ^g that equates the mean utility of a venue at the average distance with co-patron composition $(s^{\text{samerace}}, s^{\text{highinc}})$ and a venue at the average distance plus the incrementΔ^g with the average co-patron composition:

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

where $(\overline{s}^{\text{samerace}}, \overline{s}^{\text{highinc}})$ and $\overline{\text{distance}}$ denote the characteristics of the average venue. The function $\Delta^g(s^{\text{samerace}}, s^{\text{highinc}})$ is group g's willingness to travel for co-patron composition

$(s^{\text{samerace}}, s^{\text{highinc}})$.[17]

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

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

Since each parameter is g-specific, the model can be estimated separately by demographic group.

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 restaurant category. We assume that 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 exclude metro-chain pairs with very few observations and randomly sample a subset of observations from those with very many.[18]

We estimate consumer preferences using within-chain comparisons in order to distinguish consumer preferences over co-patron composition from other venue characteristics. 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 are similar because brand power and economies of scale depend on a standardized offering. In particular, the food products and quality of service typically are not a reflection of the local composition of co-patrons within each venue. In robustness checks, we focus on a subset of the

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

[18] In particular, we keep group-MSA-chain triplets that have at least 25 direct visits. When a group-MSA-chain triplet has more than 75 venues, we randomly sample only 75 venues. We randomly sample 20,000 devices from group-MSA pairs that have more than 20,000 devices and only keep group-MSA pairs with at least 200 visits. From this set, we randomly sample 750,000 visits per demographic group to reduce computational burden.

most standardized chains based on franchising terms and dispersion in establishment-level characteristics.

One may worry that co-patron composition might predict venue choice if patrons of nearby venues are demographically similar because of residential segregation. Our specification addresses this concern, because we control for bilateral distance as an individual-by-venue-specific cost shifter.[19] 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.[20] 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 exposure to high-income co-patrons will not distinguish preferences over co-patron demographics from consideration sets that vary with demographics.

4.6 Decomposition of exposure to high-income co-patrons

We decompose 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 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

[19] We could extend our venue choice model to account for residential choice as in Davis et al. (2019) Appendix C.6. This 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 their idiosyncratic preference for a specific venue).

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

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 exposure to high-income co-patrons, we use the model-predicted share of visits to each venue j by group g,

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

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 members of a demographic group.

To quantify the contributions of various mechanisms to income exposure, we report the distributions resulting from various counterfactual market shares $P_{j|g}$.[21] 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-MSA variation in demographics to exposure to high-income co-patrons, we use a uniform probability conditional on metropolitan area m, $P_{j|g} = \frac{1}{|\mathcal{J}_m|}P_{m|g}$. The difference between the nationwide uniform probability $\frac{1}{|\mathcal{J}|}$ and this measure captures the contribution of demographic differences between MSAs to income 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 residential proximity, 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 travel distance ($\delta^g = \mathbf{0}$).[22]

[21] This is similar to the approach Li (2023) uses to quantify the role of preferences versus constraints in generating residential racial segregation in the U.S. and Mayer and Puller (2008) use to quantify the role of the exogenous school environment versus preferences in the formation of social links.

[22] The last two counterfactual scenarios 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 (e.g., within-MSA to within chain-MSA), lend themselves to simple comparisons because they differ in only one respect. Non-nested scenarios, such as alternatively setting preference parameters for distance or 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.

5 Estimation results

In this section, we estimate preferences over travel distance and co-patron composition. We document notable regularities in demographic 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.

Our estimates of distance coefficients δ^g imply distance elasticities around -2.2. These distance elasticities capture both the cost of longer travel distances and the substitutability of venues within a restaurant chain. Higher-income individuals have larger 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 among all restaurants, consistent with venues within the same chain being closer substitutes.[23]

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 high-income share of co-patrons and the same-race share of co-patrons. Figures 3 and 4 present two different representations of these preferences. Figure 3 depicts preferences 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 color of the venue captures willingness to travel to that venue in co-patron composition space, $\Delta^g(s^{\text{samerace}}, s^{\text{highinc}})$ from equation (3). 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

[23] Athey et al. (2018), Davis et al. (2019), and Couture et al. (2023) estimate 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	25p	75p
Low-Income	Asian	-1.29	-0.23	-2.15	-1.59	-2.24
Low-Income	Black	-1.25	-0.23	-2.11	-1.54	-2.20
Low-Income	Hispanic	-1.26	-0.27	-2.24	-1.59	-2.35
Low-Income	White	-1.11	-0.32	-2.29	-1.51	-2.42
High-Income	Asian	-1.25	-0.26	-2.21	-1.58	-2.32
High-Income	Black	-1.11	-0.30	-2.24	-1.50	-2.36
High-Income	Hispanic	-1.22	-0.30	-2.34	-1.60	-2.46
High-Income	White	-1.00	-0.36	-2.33	-1.46	-2.48

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 Same-Race	β_4^g Quadratic Same-Race	β_5^g Interaction Term
Low-Income	Asian	4.13	-3.20	12.19	-9.93	-4.02
		(0.08)	(0.07)	(0.12)	(0.16)	(0.14)
Low-Income	Black	5.06	-4.20	4.94	-3.86	-2.18
		(0.07)	(0.05)	(0.05)	(0.04)	(0.06)
Low-Income	Hispanic	6.08	-4.21	7.08	-4.99	-4.23
		(0.07)	(0.05)	(0.08)	(0.06)	(0.07)
Low-Income	White	2.94	-4.13	2.17	-1.70	2.35
		(0.07)	(0.06)	(0.09)	(0.08)	(0.09)
High-Income	Asian	5.55	-3.21	12.78	-12.31	-2.21
		(0.07)	(0.05)	(0.08)	(0.10)	(0.09)
High-Income	Black	5.35	-3.18	4.15	-3.61	-0.29
		(0.07)	(0.05)	(0.05)	(0.04)	(0.05)
High-Income	Hispanic	7.72	-4.18	7.14	-5.10	-4.49
		(0.09)	(0.06)	(0.09)	(0.08)	(0.08)
High-Income	White	4.21	-3.65	4.47	-3.70	2.59
		(0.09)	(0.07)	(0.11)	(0.09)	(0.10)

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. Panel A reports the distance elasticity at the mean and 25$^{\text{th}}$ and 75$^{\text{th}}$ percentiles of trip distance, which are 6.4, 1.9, and 7.8 kilometers, respectively.

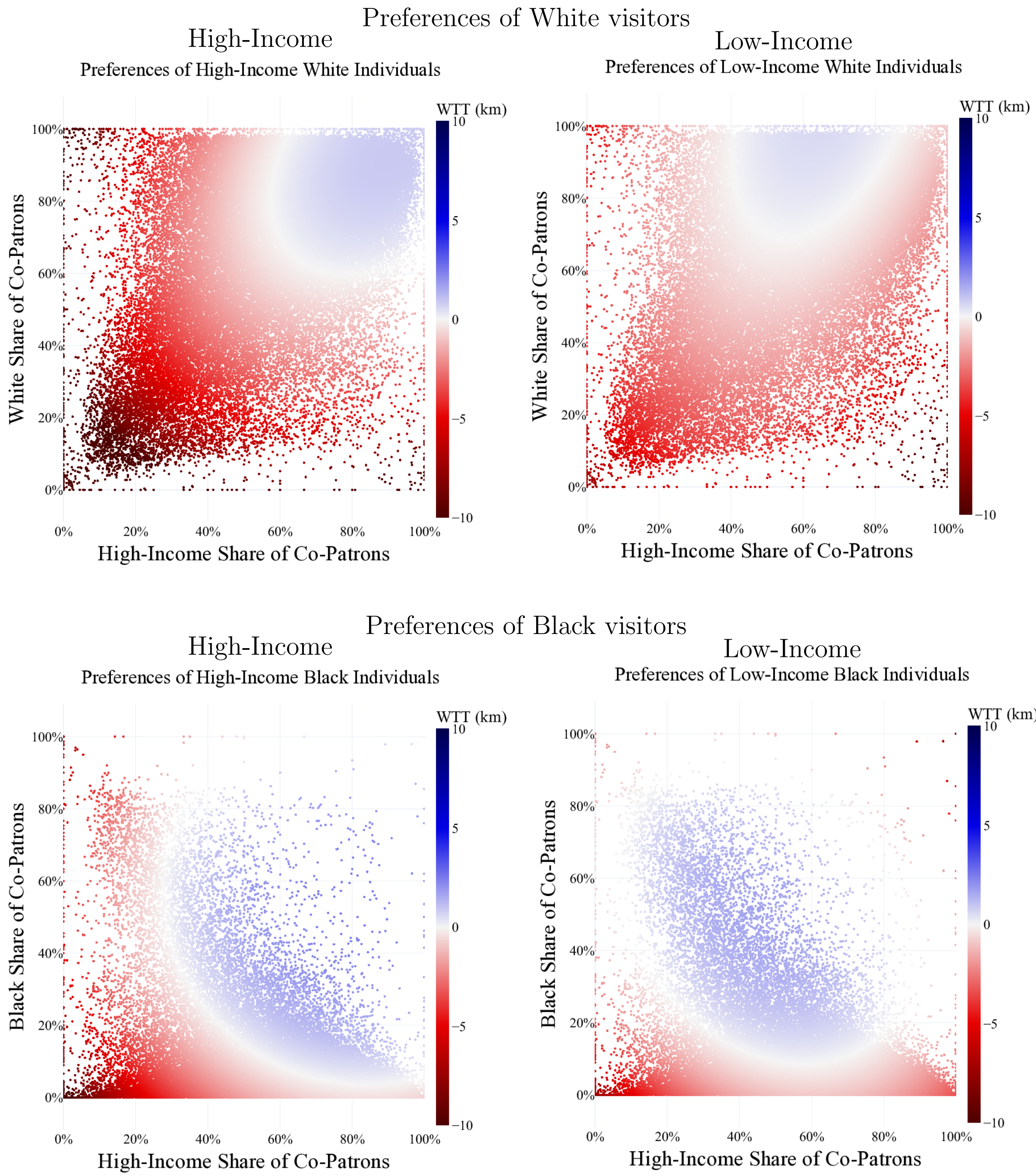
that high-income individuals have monotone preferences over high-income co-patron share, increasing from left to right in each plot, whereas their low-income counterparts have non-monotone preferences for co-patron income. Figure 4 Panel A quantifies the strength of income preferences for each group by depicting variation in willingness to travel for establishments with the median same-race share. Across all racial groups, high-income individuals are willing to travel 2.3 to 3.1 additional kilometers to visit a venue at the 95th percentile of high-income co-patron share, relative to a venue at the 5th percentile. Low-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 Asian, Black and Hispanic individuals are willing to travel around 1.4 additional kilometers to visit a venue with their most preferred income mix, while low-income White individuals have an even lower willingness to travel of around 0.7 kilometers.

Turning to preferences for same-race co-patrons, we find that all demographic groups exhibit substantial racial homophily. In all eight panels of Figure 3, the most preferred venues have high same-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 difficult because the observed same-race shares differ greatly. White individuals are the racial majority in most venues, while non-White individuals have few venues in their choice sets with large same-race shares. That said, Black, Hispanic, and White individuals have similar same-race preferences, in the sense that all are willing to travel about 2.1 km farther to visit a venue in the 95th percentile of same-race share rather than one in the 5th percentile. The racial homophily of Asian individuals appears to be stronger, albeit on very limited support. Finally, we note that racial and income preferences are roughly similar in magnitude. This similarity is important for how people trade-off income exposure for racial exposure, as we show in Section 6.

Our paper emphasizes broad patterns of demographic preferences, but our estimates offer some finer insights into what individuals prefer. Here, we briefly discuss two additional features of these preferences. First, Figure 4 shows that preferences are concave in co-patron shares. People are most willing to travel to avoid being an overwhelming racial minority or in a venue heavily patronized by low-income people. Second, we estimate significant interactions between racial and economic composition. Table 2 reports positive interaction terms for White individuals, indicating that they value high-income co-patrons more when surrounded by White co-patrons. For other racial groups, the reverse holds: the high-income share matters less as same-race share increases. This suggests that all four racial groups have sharper income preferences when co-patrons are White.[24]

[24]Seven out of eight interaction terms are between 2.2 and 4.5 in absolute value. For example, for high-

Figure 3: Preferences over co-patron demographics



Notes: Each plot visualizes the preference estimates over co-patron composition reported by equation (2) for each demographic group. Preferences are expressed in willingness to travel in kilometers relative to the average venue, $\Delta^g(s^{\text{samerace}}, s^{\text{highinc}})$, as defined in equation (3). Each dot corresponds to an individual venue. *Continues on next page.*

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

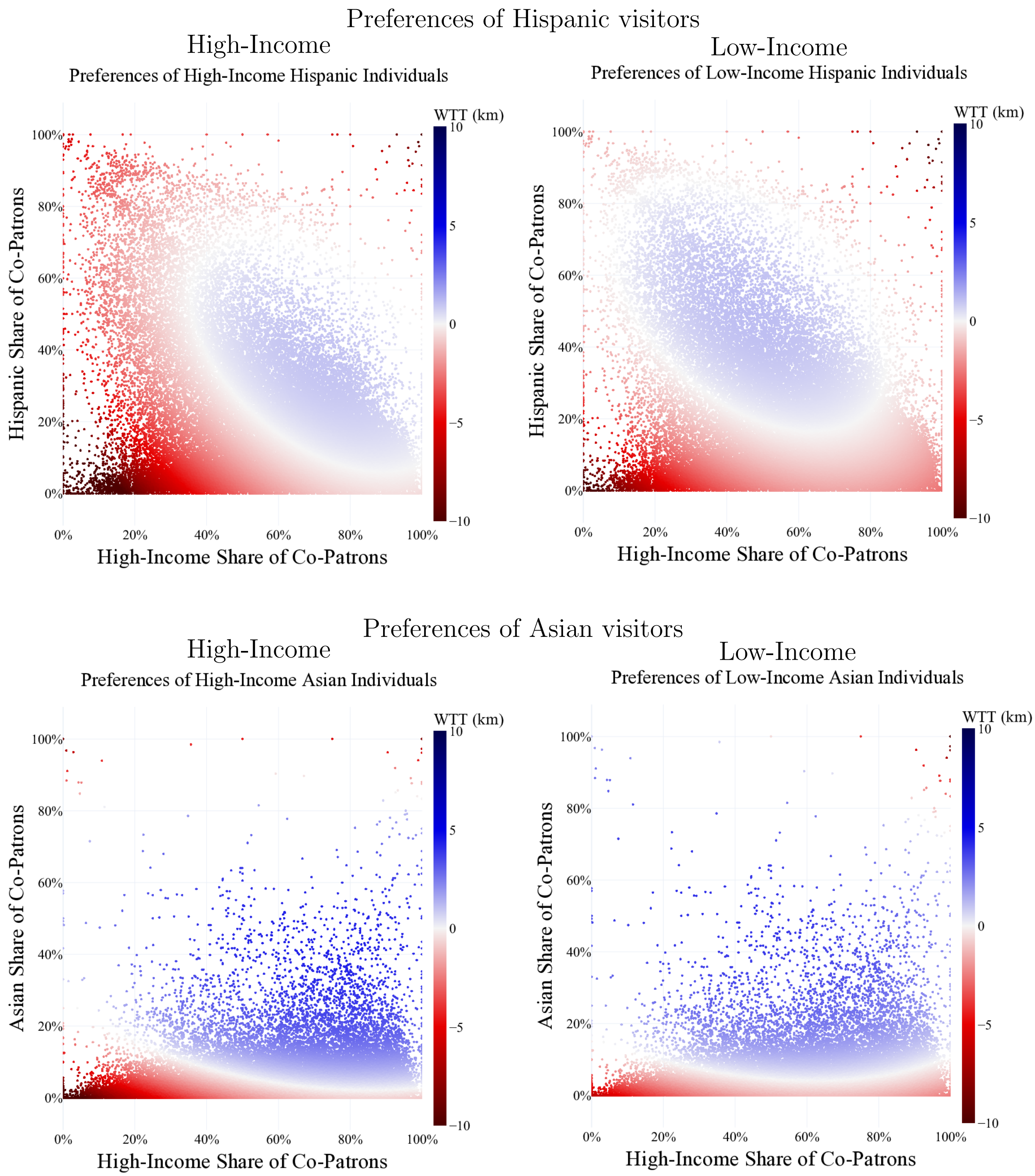
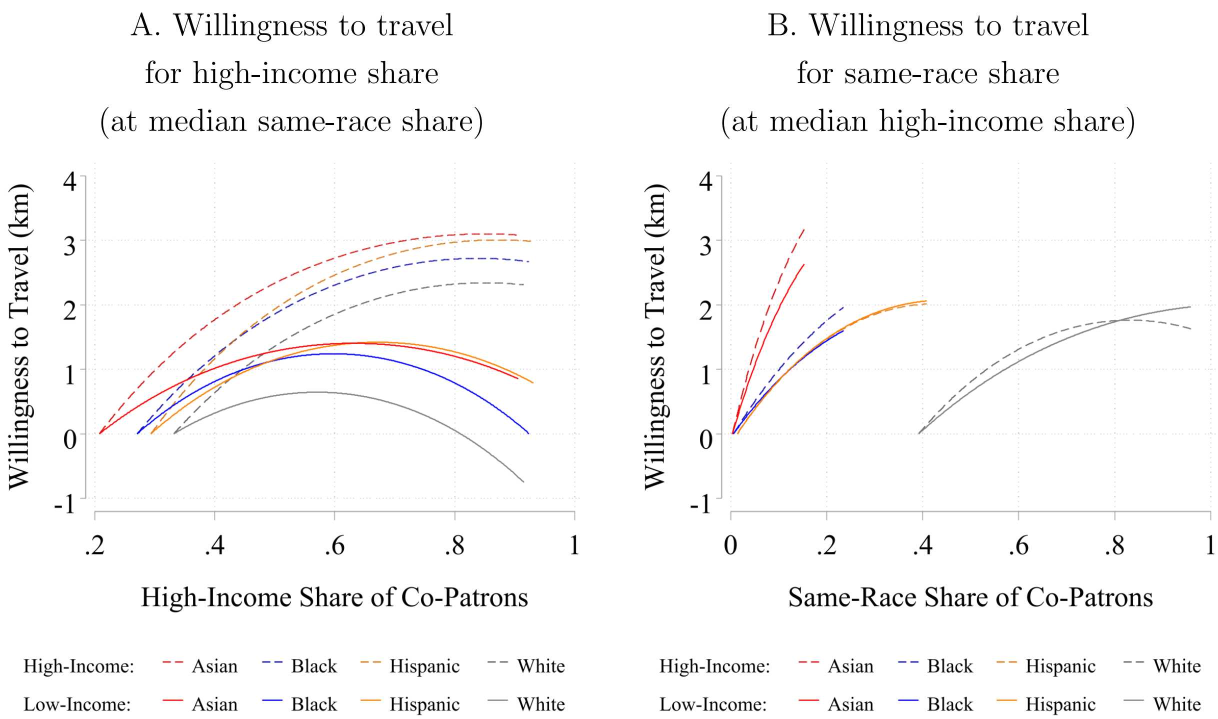


Figure 4: Preferences for high-income and same-race co-patrons



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 displays preferences over the high-income share of co-patrons. It shows willingness to travel relative to a venue at the 5th percentile of the high-income share distribution across all venues, holding same-race share fixed at its median value. Panel B displays preferences over the same-race share. It shows willingness to travel relative to a venue at the 5th percentile of the same-race share distribution across all venues, holding high-income share fixed at its median value.

Translated to dollars, these parameter estimates imply substantial willingness to pay for preferred demographic 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 95th percentile of high-income co-patron share rather than one in the 5th percentile implies a \$2.5–\$3 difference in willingness to pay per trip.[25] The 2km difference in willingness to travel between the 5th and 95th percentile of same-race share for Black, Hispanic, and White individuals translates to a \$2 difference per trip. The

income Hispanic individuals when visiting an establishment in the top quartile of Hispanic co-patron share, they are willing to travel an extra 1.73 kilometers to visit a higher-income venue (comparing venues at the 95th vs 5th percentile of high-income share). This effect is even stronger at establishments in the bottom quartile of Hispanic co-patron share, where high-income Hispanic individuals are willing to travel an additional 3.85 kilometers.

[25] Assuming 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), willingness to travel an additional kilometer is equivalent to about one dollar.

average US driver makes more than 500 consumption trips per year (Couture, Duranton, and Turner, 2018), so these estimates imply a \$1,000–1,500 annual willingness to pay to span the range of available demographic exposure in either income or race. 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, demographic preferences may therefore be an important determinant of neighborhood choice.

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 $f_1(\ln \text{distance}_{ij}; \delta^g)$. 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 $f_2(s_j^{\text{samerace}}, s_j^{\text{highinc}}; \beta^g)$ is a first-degree polynomial so there is one coefficient β_y^g for high-income share and one coefficient β_r^g for same-race share:

$$Y_{ij} = \delta_1^g \ln \text{distance}_{ij} + \delta_2^g \left(\ln \text{distance}_{ij}\right)^2 + \beta_y^g s_j^{\text{highinc}} + \beta_r^g s_j^{\text{samerace}}. \tag{7}$$

Broadly, these robustness checks deliver preference estimates that are quite similar across the various samples and specifications.

5.2.1 Within-chain heterogeneity

We select the most standardized chains based on two characteristics. The first is the coefficient of variation in the Google Places star rating across venues within the chain.[26] Less variation in reviewer ratings across venues suggests a more standardized service. The second characteristic is ownership structure. Following Williamson (1991)'s argument that franchis-

[26]The Google Places data on restaurant venue location and characteristics comes from Akbar et al. (2023). We were able to match 41 percent of Precisely PlaceIQ venues in our estimation sample to a venue in that Google Places data. Appendix A.6 provides more details on the Google Places data and variable construction.

ing facilitates local adaptation, we expect chains with franchisees to be less standardized than owner-operated chains.[27] 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."[28]

These two measures of chain standardization are consistent with one another. The Google Places star rating variation of a franchised chain is on average twice as 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 among the 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 among the most standardized using these metrics.[29]

Figure B.5 reports estimation results for all eight demographic groups within four 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 in star rating. 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 across venues in product attributes and service quality.

Our second set of robustness checks addresses within-chain heterogeneity by controlling for more venue and neighborhood characteristics. Figure B.6 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 Precisely PlaceIQ.[30] Adding these covariates has little impact on estimated preferences for same-race co-patrons, but it raises our estimates of preference for high-income co-patrons for all groups. Finally, we control for the demographic

[27] 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.

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

[29] McDonalds, Subway, Starbucks, and Burger King are all in the top third of chains for star rating variation, 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.

[30] We include variation in square footage, but these data are often mis-measured (it sometimes includes parking lots for instance) and we are not confident that it captures true variation across chains.

composition of the residents of or visitors to the neighborhood in which the venue is located.[31] 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 same-race and high-income residents in the venue's census tract or the shares of same-race and high-income co-patrons within all other commercial venues located in a venue's census tract (Figure B.7).[32]

5.2.2 Travel-cost specification

The estimates reported in Section 5.1 could be biased by some component of travel costs that is not captured 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 than driving. Or individuals may have a particular 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 B.8 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.[33]

Figure B.9 reports preference estimates using three alternative functions of distance. The first uses a linear function of log distance, and the second uses a cubic polynomial of log distance, which is more flexible than our baseline quadratic polynomial. The third 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.

[31] The correlations between a venue's co-patron shares and the demographic shares of the block group in which it is located are quite high: 0.58 (high income), 0.63 (Asian), 0.75 (Black), 0.81 (Hispanic), and 0.77 (White). If we were to use only neighborhood-level residential demographics, as Davis et al. (2019) do, we would estimate similar but smaller coefficients than those on the venue-level co-patron demographics. When we include both demographic covariates, the coefficients on the venue-level shares are substantially larger than those on the block-group-level shares, suggesting that individuals care primarily about venue composition, not neighborhood composition.

[32] One exception is the income preferences of low-income individuals, which weaken 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.

[33] These MSA-level statistics come from the 2017 National Household Travel Survey (NHTS). Almost 90% of trips to commercial venues in the United States are by car. Appendix A.7 provides more detail on the NHTS data and variable construction.

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 6 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, restaurants. Restaurants and coffee shops have been singled out as a plausible setting for demographic 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 demographic preferences within other kinds of commercial venues that have chains. A priori, it is unclear whether to expect weaker or stronger preference estimates in other settings. For instance, preferences for co-patrons within retail chains may be weaker if demographic exposure is less salient in that 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 shelf space for Asian food in a predominantly Asian neighborhood. Figure B.10 reports estimated preferences for each demographic group for eight distinct venue categories: banks, big-box stores, convenience stores & gas stations, grocery stores, gyms, pharmacies, restaurants, and all business categories pooled together.[34] We find some variation in the magnitude of preferences across categories, but the results confirm that the preference patterns documented in Section 5.1 are not unique to restaurants. For all business categories, individual exhibit racial homophily and prefer venues with more high-income co-patrons, 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.

[34] Table A.4 shows the five largest chains in each category.

6 Determinants of demographic exposure

This section examines how preferences over co-patron demographics contribute to realized exposure to high-income individuals. The preferences estimated in Section 5 show that, conditional on where they live, people are willing to travel significant distances for different co-patrons demographics. Section 6.1 shows that these choices impact income segregation directly. For example, high-income Asian, Hispanic, and White individuals' preferences for high-income co-patrons are sufficient to single-handedly explain most of their exposure to high-income co-patrons. Beyond their direct effects, preferences over co-patron demographics are also informative about residential sorting. Section 6.2 shows that, within demographic groups, individuals reside in and move to neighborhoods with demographics that align with their preferences over co-patron demographics.

6.1 Decomposition of exposure to high-income co-patrons

Table 1 documented large differences across demographic groups in their exposure to high-income co-patrons. To quantify the importance of various factors in explaining these disparities, we compute model-predicted visits under the different counterfactual scenarios described in Section 4.6.

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

		Low Income				High Income			
		Asian (1)	Black (2)	Hispanic (3)	White (4)	Asian (5)	Black (6)	Hispanic (7)	White (8)
(1)	Uniform within MSA	-0.02	-0.01	-0.05	-0.01	0.01	0.01	-0.02	0.01
(2)	Uniform within MSA-chain pair	-0.02	-0.02	-0.06	-0.00	0.04	0.02	-0.01	0.04
Within MSA-chain nests									
(3)	Residential proximity	-0.06	-0.16	-0.15	-0.03	0.11	-0.01	0.04	0.11
(4)	Demographic preferences only	-0.00	-0.05	-0.07	0.01	0.09	0.04	0.03	0.10
(5)	Income preferences only	-0.00	-0.01	-0.02	-0.00	0.09	0.06	0.06	0.09
(6)	Race preferences only	-0.01	-0.04	-0.06	0.01	0.05	-0.01	-0.01	0.06
(7)	Model-predicted visits	-0.06	-0.16	-0.15	-0.03	0.14	-0.00	0.05	0.13
(8)	Estimation-sample visits	-0.06	-0.16	-0.15	-0.02	0.14	-0.00	0.05	0.14

Notes: This table shows the average high-income share of co-patrons an individual of each group would be exposed to under different counterfactual visit scenarios. All rows are differenced from a "uniform" counterfactual scenario in which devices visit venues with uniform probability nationwide. The first row reports the counterfactual scenario in which devices visit venues uniformly within their MSA of residence. while the second row reports the scenario with uniform visits within their MSA of residence and choice of chain. The third row reports the counterfactual scenario in which devices consider their distance dis-utility while ignoring preferences for co-patron characteristics. The fourth row reports the counterfactual scenario in which devices consider their preferences for co-patron characteristics while ignoring their distance dis-utility. The fifth row reports the counterfactual scenario in which devices only consider their preferences for high-income co-patrons. The sixth row reports the counterfactual scenario in which devices only consider their preferences for same-race co-patrons. The seventh row reports the counterfactual scenario in which devices consider both their preferences for co-patron compositions and distance dis-utility. Finally, the eighth row shows the actual exposure based on the home-venue-home visits in our estimation sample.

Table 3 reports how the mean exposure to high-income co-patrons varies across these counterfactuals. Each cell describes the visit-weighted average high-income share of co-patrons that a given demographic group experiences in a given counterfactual scenario. All rows show changes in mean exposure relative to a "uniform benchmark": the exposure each group would receive if all venues were visited with uniform probability. Our venue-choice model fits the data well, as the model-predicted visits generate exposures very close to those in the estimation sample and all observed visits to chain restaurants.

The other rows of Table 3 introduce different counterfactual scenarios. The "uniform within MSA" and "uniform within MSA-chain pair" rows reveal the importance of (unmodeled) variation across MSA-chain nests for explaining differences in exposure. Then, we compare how the two within-nest factors included in our model of venue choice—residential proximity and demographic preferences—determine model-predicted visits.

The "uniform within MSA" row depicts the change in mean exposure to high-income co-patrons if individuals visited venues uniformly within their MSA of residence, relative to a national uniform exposure benchmark. Accounting for metropolitan differences barely shifts exposure relative to the national average benchmark. The only substantial shift is for low-income Hispanic individuals, who tend to reside in poorer cities. For them, differences across MSAs explain one-third of their experienced income exposure relative to the benchmark (-0.05 of -0.15).

The "uniform within MSA-chain pair" row reports the change in mean exposure if individuals visited venues uniformly within the chains that they visit in the MSA where they reside, again relative to the benchmark of national uniform high-income exposure. Accounting for chains has little effect: most exposure differences arise within MSA-chain pairs. The largest shifts in exposure are among high-income Asian and White 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. While there are statistically significant differences in the types of chains visited by different groups—for instance, high-income individuals are more likely to visit Starbucks than Dunkin' Donuts—these differences are not large enough to meaningfully impact experienced income segregation.[35] In summary, differences in the distribution of groups across cities and chains fail to explain most of the differences in income exposure across groups.

Thus, most of these differences in income exposure must be explained by proximity to venues or demographic preferences. The "residential proximity" and "demographic preferences only" rows of Table 3 summarize the contribution of each factor to predicted income exposure. Specifically, the "residential proximity" row reports the income exposure outcomes that would result if individuals chose which of a chain's venue to visit based only on distance parameters δ^g absent any co-patron preferences ($\beta^g = \mathbf{0}$). Conversely, the "demographic preferences only" row reports the income exposure outcomes that would result if individuals chose which of a chain's venue to visit based only on preferences for co-patron composition absent any disutility of distance ($\delta^g = \mathbf{0}$).

Residential distance dramatically reduces the high-income exposure of low-income individuals from all racial groups. The neighborhoods in which low-income individuals reside are the most important factor in explaining why they are less exposed to high-income co-patrons. In fact, the "residential proximity" and the "model-predicted visits" rows of columns 1 through 4 in Table 3 are nearly identical. In this sense, the income segregation

[35] Chain choice depends on tastes for chain product offerings, preferences for average co-patron composition within each chain, and proximity to chain venues. Given the small importance of chain choice, however, we conclude that most of demographic exposure is determined by venue choice within, not across, chains. Nilforoshan et al. (2023) suggest that the availability of venues catering to richer people is the main reason why larger cities feature more experienced segregation. We find that this hypothesis does not explain differences in income exposure across demographic groups, because these demographic differences persist *within* business chains.

of consumption venues reflects the income segregation of residences. For high-income individuals, however, residential proximity increases high-income exposure, because they live in neighborhoods near high-income venues. The exception is high-income Black individuals, who live in much poorer neighborhoods than their high-income counterparts in other racial groups. For them, residential proximity diminishes income exposure relative to a uniform visit benchmark.

Demographic preferences also play an important role in explaining racial differences in income exposure. In particular, racial homophily impacts income exposure differently across racial groups. As an example, Black and White individuals share similar preferences for high-income co-patrons and similar levels of racial homophily, but these same preferences drive Black individuals to visit venues with much smaller shares of high-income co-patrons. This is because Black and Hispanic individuals, who have lower income on average, must choose between visiting high-income venues or heavily same-race venues.[36] Asian and White individuals do not face this trade-off, and eliminating same-race preferences, in row 5 of Table 3, considerably narrows the income exposure gap between racial groups.

Overall, residential proximity dominates demographic preferences for low-income individuals, whose experienced income exposure is lower than what their demographic preferences alone would suggest. For high-income Asian, Hispanic and White individuals, however, either residential proximity or demographic preferences suffice to account for much of their observed high-income exposure. These groups live in neighborhoods where venues that suit their demographic preferences—those with large shares of high-income and same-race co-patrons—are located nearby. These results are consistent with either individuals choosing neighborhoods based on their demographic preferences or with residential experiences in those neighborhoods affecting preferences. The next section investigates this further.

6.2 Residential sorting on demographic preferences

We now use our model of venue choice to study the link between demographic preferences and residential sorting. We perform two exercises. First, we estimate the model for individuals in the same demographic group who reside in neighborhoods with different demographic mixes. We find that individuals are sorted across neighborhoods in a way that is correlated with their demographic preferences. Second, we study the evolution of demographic preferences around residential moves. When individuals move, they move to neighborhoods that suit their pre-move demographic preferences, and, in turn, their demographic preferences partially converge to those of their new neighborhood over the next five months.

[36] Bayer and McMillan (2005) suggest that the scarcity of high-income majority Black neighborhoods may explain why Black households live in poorer neighborhoods than White households with similar incomes.

6.2.1 The extent of sorting on demographic preferences

Do people living in neighborhoods with more high-income or same-race residents have stronger preferences for high-income or same-race co-patrons? We divide Census tracts into income terciles and same-race terciles based on the composition of their residents. Leveraging our building-level demographic data, we estimate the model of equation (2) separately for devices residing in each income tercile and race tercile.[37] The results reveal preference heterogeneity within demographic groups. Figure 5 depicts these preference estimates for high- and low-income White individuals (dashed and solid lines, respectively).[38]

Figure 5 shows that White individuals' preferences over co-patron demographics are aligned with their neighborhoods' demographics. The upper-left panel shows that White residents of higher-income neighborhoods have stronger preferences for high-income co-patrons. This is true for both high- and low-income individuals. These differences in preferences across terciles of neighborhood income are large: a low-income resident of a top tercile neighborhood (solid red) has preferences for co-patron income similar to a high-income resident of a middle tercile neighborhood (dashed gray).

The lower-right panel of Figure 5 shows a similar alignment of same-race preferences with neighborhood demographics: White individuals who reside in heavily White neighborhoods exhibit stronger preferences for same-race co-patrons. Those residing in the top tercile of neighborhoods with the highest White share would travel five times farther than those in the bottom tercile to visit a venue in the 95th percentile of White co-patron share rather than a venue at the 5th percentile.

Overall, we find that within race-income groups, residential demographics and preferences over co-patron demographics are meaningfully correlated. These patterns could arise because individuals select neighborhoods with residents matching their preferred co-patron composition or because residential experiences affect those preferences (e.g., the intergroup contact hypothesis). The movers analysis in the next subsection speaks to these links.

6.2.2 Origins of sorting on demographic preferences

To study the relationship between residential demographics and preferences for venue co-patrons, we estimate preference parameters at a monthly frequency before and after moves. To achieve sufficient sample sizes, we study high-income White individuals who move between

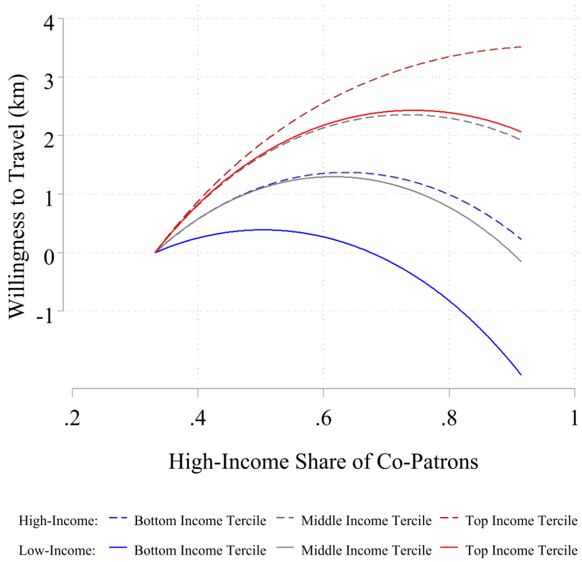
[37]This design requires building-level demographic data. We could not observe variation in preferences across neighborhoods within demographic groups if these groups were defined using neighborhood characteristics.

[38]Analogous figures for other racial groups are in Appendix B.1. They show generally similar patterns, but are noisier due to smaller samples. For example, few low-income Black individuals live in the upper tercile of the tract income distribution.

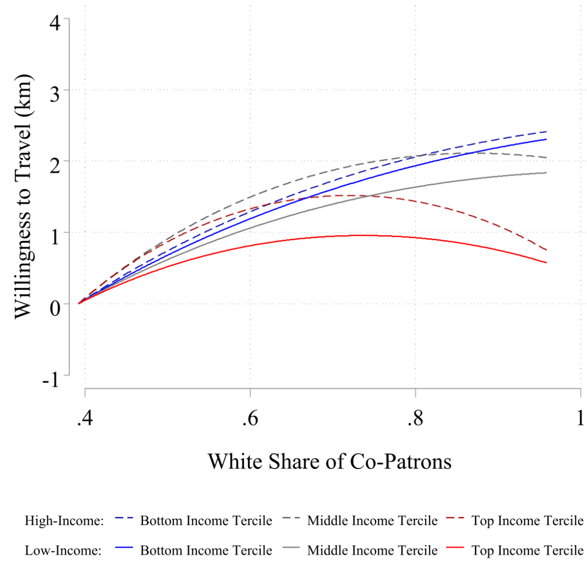
Figure 5: Preference heterogeneity across White visitors by residential neighborhoods

Panel A: Preferences by residential neighborhood income

Willingness to travel for high-income share

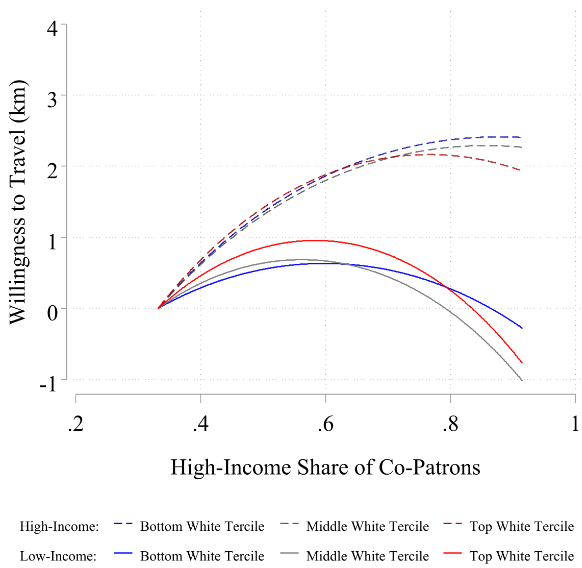


Willingness to travel for White Share

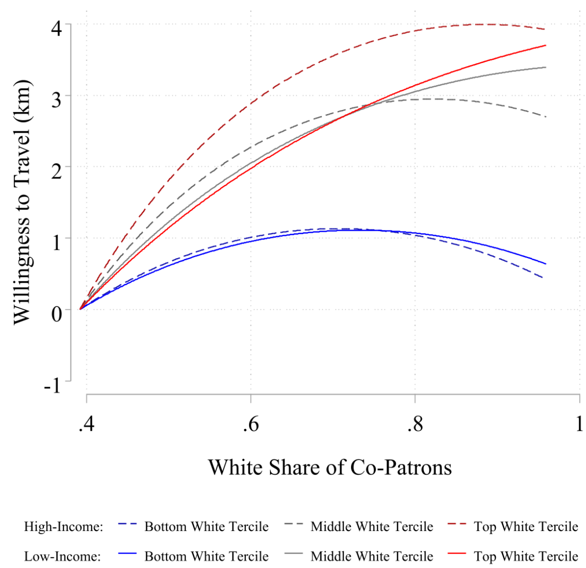


Panel B: Preferences by residential neighborhood White share

Willingness to travel for high-income share



Willingness to travel for White share



Notes: These plots are analogous to those in Figure 4 except that each line depicts preferences for high- or low-income White visitors who reside in neighborhoods (Census tracts) with different demographics. Residential tract high-income terciles are defined using tract-level high-income share weighted by high-income tract population. Similarly for same-race terciles. We estimate preferences for a randomly selected 100,000 visits by devices in a given residential tercile. To compare across neighborhood 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 5th percentile of the high-income share distribution across all venues, holding same-race share fixed at its median value. On the right, we show willingness to travel relative to a venue at the 5th percentile of the same-race share distribution across all venues, holding high-income share fixed at its median value.

MSAs for five months before and five months after their change in residence.[39] We examine how preference parameters evolve when an individual moves from an origin neighborhood in same-race tercile o to a destination neighborhood in same-race tercile d. We focus on moves across same-race terciles because moves across income terciles, shown in Appendix B.2, may arise from income shocks that could affect preferences. We estimate the parsimonious linear specification of our model—used for all robustness exercises in Section 5.2—in which preferences over co-patron composition depend on only the same-race share of co-patrons and the high-income share of co-patrons, but allow preference coefficients over both travel costs ($\delta_k^{g,od}$) and co-patron composition ($\beta_k^{g,od}$) to vary depending on the origin and destination same-race terciles (o and d) and the month relative to the move (k). Specifically, we re-estimate the model outlined in Section 4.2, replacing the utility component that varies across venues within chain (Y_{ij}) with a time-varying counterpart:

$$Y_{ijt} = \sum_{k=-5}^{5} \mathbf{1}\{t=k\} \left(\delta_{1k}^{g,od} \ln \text{distance}_{ij} + \delta_{2k}^{g,od} \ln \text{distance}_{ij}^2 + \beta_{rk}^{g,od} s_j^{\text{samerace}} + \beta_{yk}^{g,od} s_j^{\text{highinc}} \right), \tag{8}$$

where $\beta_{rk}^{g,od}$ and $\beta_{yk}^{g,od}$ are the preference coefficients on same-race share and high-income share, respectively, for group g in month k when moving from tercile o to tercile d.

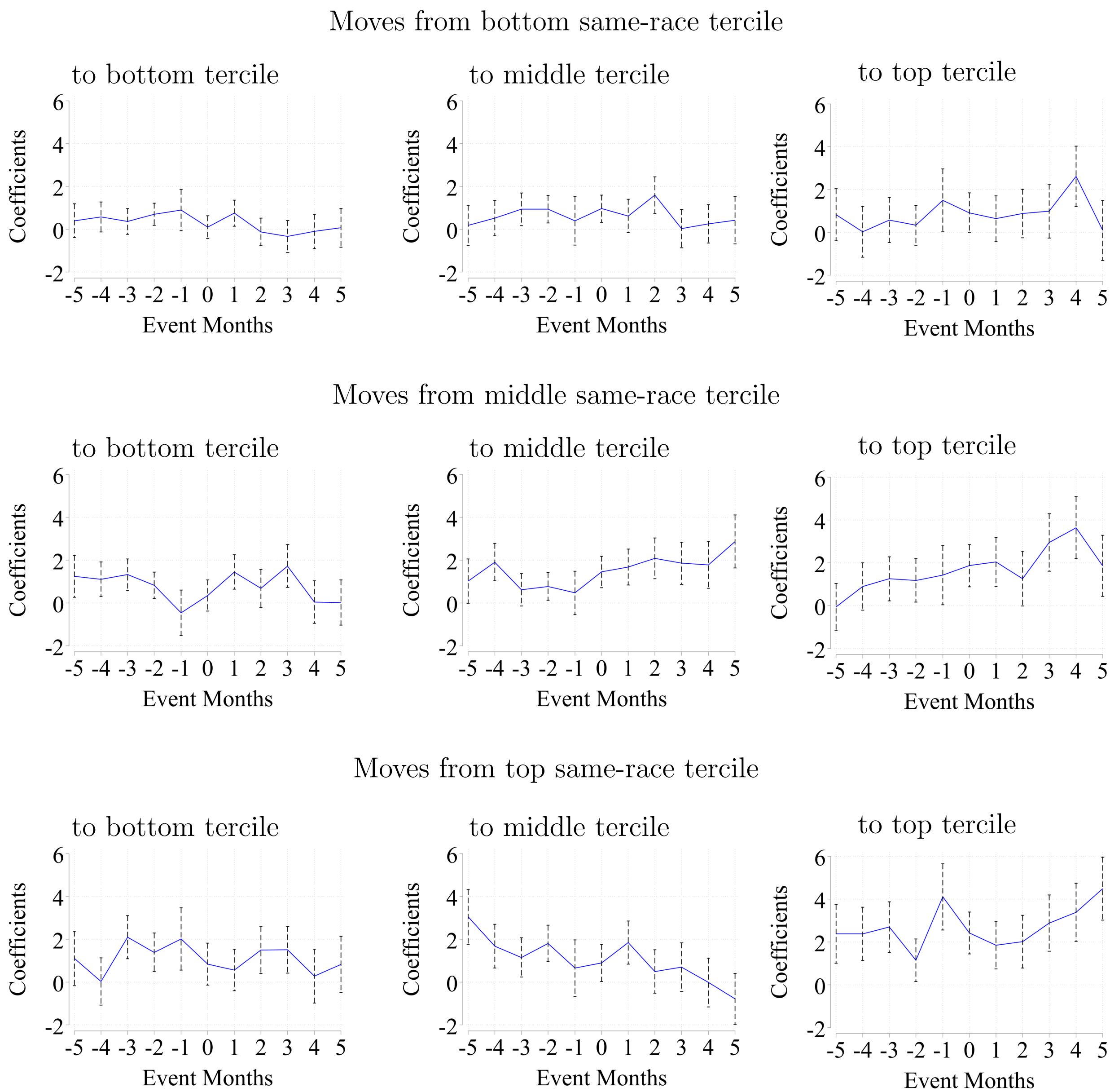
Table 4: Average same-race preference of high-income White devices before & after moves

	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
	(0.17)	(0.20)	(0.27)	(0.19)	(0.20)	(0.26)	(0.27)	(0.25)	(0.29)
Post-Move	0.06	0.65	1.03	0.71	1.96	2.27	0.92	0.53	2.84
	(0.15)	(0.18)	(0.25)	(0.19)	(0.21)	(0.27)	(0.23)	(0.22)	(0.26)

NOTES: This table reports pooled estimates of $\beta_{rk}^{g,od}$ from equation (8). For each origin-destination pair, coefficients are pooled for event-months prior to move ($t=-5$ to $t=-1$), and post-move ($t=0$ to $t=5$). The estimation samples contain home-venue-home visits to restaurants by high-income White individuals who move between MSAs, split by the same-race tercile of the origin residence (with O1 being the origin tercile with the lowest share White) and the same-race tercile of the destination residence (with D1 being the destination tercile with the lowest share White).

[39] We focus on between-MSA 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-MSA movers.

Figure 6: Preferences over same-race share by month by tercile-to-tercile move



NOTES: Each plot in this figure depicts event-month-specific, origin-race-tercile-destination-race-tercile-specific estimates of preference for same-race co-patrons $\beta_{rk}^{g,od}$ in equation (8). Each dot depicts a point estimate and the bands depict 95% confidence intervals. The estimation samples contain home-venue-home visits to restaurants by high-income White individuals who move between MSAs, split by the same-race tercile of the origin residence and the same-race tercile of the destination residence.

Figure 6 shows estimated preferences for same-race exposure before and after moves across terciles of the neighborhood same-race distribution.[40] Given the small sample 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.

Figure 6 and Table 4 reveal three patterns of interest. First, we find some evidence that individuals move to neighborhoods with demographics that suit their pre-move demographic preferences. Conditional on the demographic tercile of the origin, individuals with stronger pre-move racial preferences tend to move to destination neighborhoods with higher same-race shares. These differences are not always statistically significant, but they are especially large – a near doubling in the strength of racial homophily – when contrasting individuals moving from a third-tercile origin to a third-tercile destination (O3 to D3) with individuals moving from a third-tercile origin to a first-tercile destination (O3 to D1). In other words, high-income White individuals moving from a high- to a low-share White neighborhood have substantially weaker same-race preferences before the move.

Second, estimated preference coefficients do not jump discontinuously when individuals move. The nine panels of Figure 6 depict event studies for moves between the nine pairs of terciles of the same-race neighborhood distribution. The top-right (O1 to D3) and bottom-left (O3 to D1) plots show the greatest contrasts between origin and destination terciles. In all cases, there are neither pre-trends in preferences prior to a move nor noticeable jumps in preferences right after moving.

Third, demographic preferences partially converge to the local demographic mix after a move. The changes in estimated preference parameters, while not always statistically significant, are consistent with neighborhoods shaping preferences. Individuals who move to neighborhoods with a higher same-race share tend to exhibit stronger racial homophily after moving. For instance, the same-race preferences of movers from O1 to D3 rise by almost 60% post-move, but remain weaker than the preferences of those already living in O3. Similarly, the preferences of individuals who move to a neighborhood with a lower same-race share evolve to exhibit weaker racial homophily. Given the limited time window, we

[40] Table B.2 replicates Table 4 for income preferences following moves across neighborhood income terciles. It reports substantial increases in income preferences following moves to higher income terciles, providing additional evidence that preferences change in line with the demographic of one's new neighborhood. As expected, we do not find that individuals sort across income terciles due to their income preferences. Such moves may be explained by income shocks. Appendix B.2 reports the full set of event studies for income exposure around moves to different same-race terciles, and for same-race and income exposure around moves to different income terciles. Out of these 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.

cannot assess whether preferences fully converge, but the point estimates constitute evidence for neighborhoods shaping preferences.

Overall, our examination of movers suggests that residential sorting explains some of the documented spatial heterogeneity in demographic preferences and that exposure to new neighbors may in turn shape these preferences. Sorting appears relevant because preferences before a move predict the dominant demographic of the destination neighborhood. Residential experiences appear relevant because after a move preferences for the dominant demographic of the destination neighborhood strengthen. These results, while limited by sample sizes and time spans, would be consistent with extended contact with one's neighbors early in life influencing preferences for demographic exposure, which then evolve slowly with new experiences.

7 Conclusion

Our study offers new insights into the demographic fragmentation of American life. We measure exposure to high-income co-patrons in shared commercial spaces by combining data from millions of smartphones with building-level demographic characteristics. We estimate preferences for the racial and income composition of these spaces by studying visits to large chains in which variation in venue attributes is limited.

We find large disparities across groups in their exposure to high-income co-patrons. Black, Hispanic, and low-income individuals experience lower high-income exposure. Demographic preferences, however, are broadly shared across groups and economically large. For instance, racial homophily translates into a willingness to travel two additional kilometers to visit a venue at the 95th percentile of the same-race distribution rather than the 5th percentile, and these preferences do not attenuate at higher incomes.

Our findings have important implications for the future of American segregation. Since demographic preferences alone can explain a substantial portion of income segregation in shared spaces, eliminating differences in product tastes and residential proximity may not suffice to close the gaps in exposure to high-income individuals across demographic groups. Moreover, if these demographic preferences influence behavior in other settings, such as neighborhoods and schools, removing structural barriers alone is unlikely to fully integrate American society.

Our analysis, however, reveals settings in which demographic preferences are less pronounced. In particular, preferences are weaker in more integrated neighborhoods and weaken following a move to a more integrated area. Consistent with the intergroup contact hypothesis, integrated spaces may therefore promote further integration over time by attenuating demographic preferences.

References

Akbar, Prottoy A., Victor Couture, Gilles Duranton, and Adam Storeygard. 2023. "The Fast, the Slow, and the Congested: 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.

Allport, G.W. 1954. *The Nature of Prejudice. The Nature of Prejudice*, Addison-Wesley Publishing Company.

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."

Backstrom, Jesse, and Richard T Woodward. 2021. "Racial Distance Matters Too: Racial and Ethnic Reference Dependence in Recreation Decision Making." *Available at SSRN 3818265.*

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, Fernando Ferreira, and Robert McMillan. 2007. "A unified framework for measuring preferences for schools and neighborhoods." *Journal of Political Economy*, 115(4): 588–638.

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.

Bruch, Elizabeth E. 2014. "How population structure shapes neighborhood segregation." *American Journal of Sociology*, 119(5): 1221–1278.

Bursztyn, Leonardo, Thomas Chaney, Tarek A Hassan, and Aakaash Rao. 2024. "The immigrant next door." *American Economic Review*, 114(2): 348–384.

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."

Cantoni, Enrico, and Vincent Pons. 2022. "Does context outweigh individual characteristics in driving voting behavior? Evidence from relocations within the United States." *American Economic Review*, 112(4): 1226–1272.

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.

Choi, Jinkyong, Jorge Guzman, and Mario L Small. 2024. "Third Places and Neighborhood Entrepreneurship: Evidence from Starbucks Cafés." National Bureau of Economic Research.

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.

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

Lowe, Matt. 2021. "Types of Contact: A Field Experiment on Collaborative and Adversarial Caste Integration." *American Economic Review*, 111(6): 1807–44.

Magontier, Pierre, Markus Stefan Schläpfer, and Maximilian von Ehrlich. 2022. "The Fragility of Urban Social Networks-Mobility as a City Glue."

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

Mayer, Adalbert, and Steven L Puller. 2008. "The old boy (and girl) network: Social network formation on university campuses." *Journal of public economics*, 92(1-2): 329–347.

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.

Nilforoshan, Hamed, Wenli Looi, Emma Pierson, Blanca Villanueva, Nic Fishman, Yiling Chen, John Sholar, Beth Redbird, David Grusky, and Jure Leskovec. 2023. "Human mobility networks reveal increased segregation in large cities." *Nature*.

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.

USDOT, Federal Highway Administration. 2017. "2017 National Household Travel Survey."

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.

Yabe, Takahiro, Bernardo García Bulle Bueno, Xiaowen Dong, Alex Pentland, and Esteban Moro. 2023. "Behavioral changes during the COVID-19 pandemic decreased income diversity of urban encounters." *Nature Communications*, 14(2310).

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. Precisely 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 Precisely PlaceIQ based on their experience correlating their data to a diverse array of truth sets, including consumer spending data and foot-traffic counts. Precisely PlaceIQ also reports a lower bound for the visit's duration based on the time between consecutive pings at the same venue.

When two venues are in close geographic proximity, a single visit may have an attribution score exceeding Precisely PlaceIQ's recommended threshold value for multiple venues. Following the methodology outlined in Couture et al. (2021), in these cases, 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. If both polygons have identified business categories or neither have identified business categories we drop those visits.

We include all visits 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 the 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

[41] The most common case of such an overlap happens when the basemap contains one polygon representing a business establishment and a second polygon representing both that building and the accompanying parking lot.

Precisely PlaceIQ data. This allows us to construct a weekly panel of home locations for a subset of devices using the following assignment methodology:

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 that device 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, then we keep the home that appears for the most weeks.

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

Precisely PlaceIQ provides us with demographic data at the building level for around 36 million residential buildings. This includes information on standard demographic information

[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 where 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.1: Homogeneous Buildings by Race and Income

Category	Group	Buildings	Percent
Race/Ethnicity	All	34,547,538	100
	Asian	1,013,998	3
	Black	2,198,715	6
	Hispanic	3,936,421	11
	White	24,203,483	70
Income	All	34,547,538	100
	Low Income	14,369,436	42
	High Income	19,839,966	57

NOTES: This table shows the number of buildings for we which we have information on race/ethnicity and income. The "All" rows show that number for all buildings, and other rows show the number of buildings that are "nearly homogeneous" ($> 67\%$) for the four racial groups and two income groups.

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 ($< \$75,000$, the bracket cutoff closest to the national median in 2019), and high-income ($> \$75,000$). For racial/ethnic categories, we aggregate the provided bins to non-Hispanic Asian, non-Hispanic Black, non-Hispanic White, Hispanic, and all other racial/ethnic groups not covered by our study.[44]

Table A.1 shows the number of buildings that contain 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 two income groups. 99% of buildings are at least 67% low- or high-income. 91% of buildings are at least 67% Asian, Black, Hispanic, or White.

A.4 Building-level data representativeness

In the main text, we show that the smartphone sample for which we have building-level data is spatially representative. For instance, within a county, we have about the same number of devices in block groups with a high White share as in block groups with a low White share. In this appendix, we show that the building-level demographic data are highly correlated with publicly available Census demographic data when aggregated to larger spatial units. This exercise shows, for example, that we have more high-income devices in the building-level

[44] Specifically, Asian includes "Central and Southwest Asian", "Far Eastern", "Southeast Asian", Black includes "African American", White includes "Eastern European", "Jewish", "Mediterranean", "Scandinavian", "Western European", Hispanic includes "Hispanic", and the remaining groups not covered by our study are "Middle Eastern", "Native American", "Other" and "Polynesian".

data relative to the census. 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. The two county-level measures are highly correlated: the R^2 exceeds 0.81 for all four demographic shares. Given Panel B of Figure 1 showing that the smartphone sample with building demographics is broadly spatially representative across block groups within counties, the gaps between observations and the 45-degree line in Figure A.1 largely reflect differences within, rather than across, block groups.

Overall, we find that 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 regressing the high-income household share in the Census on the same share in the Precisely PlaceIQ building-level data yields a coefficient of 1.36 and an R^2 of 0.84. Similar regressions using share of Black, Hispanic, and White households yield coefficients of 0.84, 0.87, and 0.89, and R^2s of 0.81, 0.97, and 0.92.

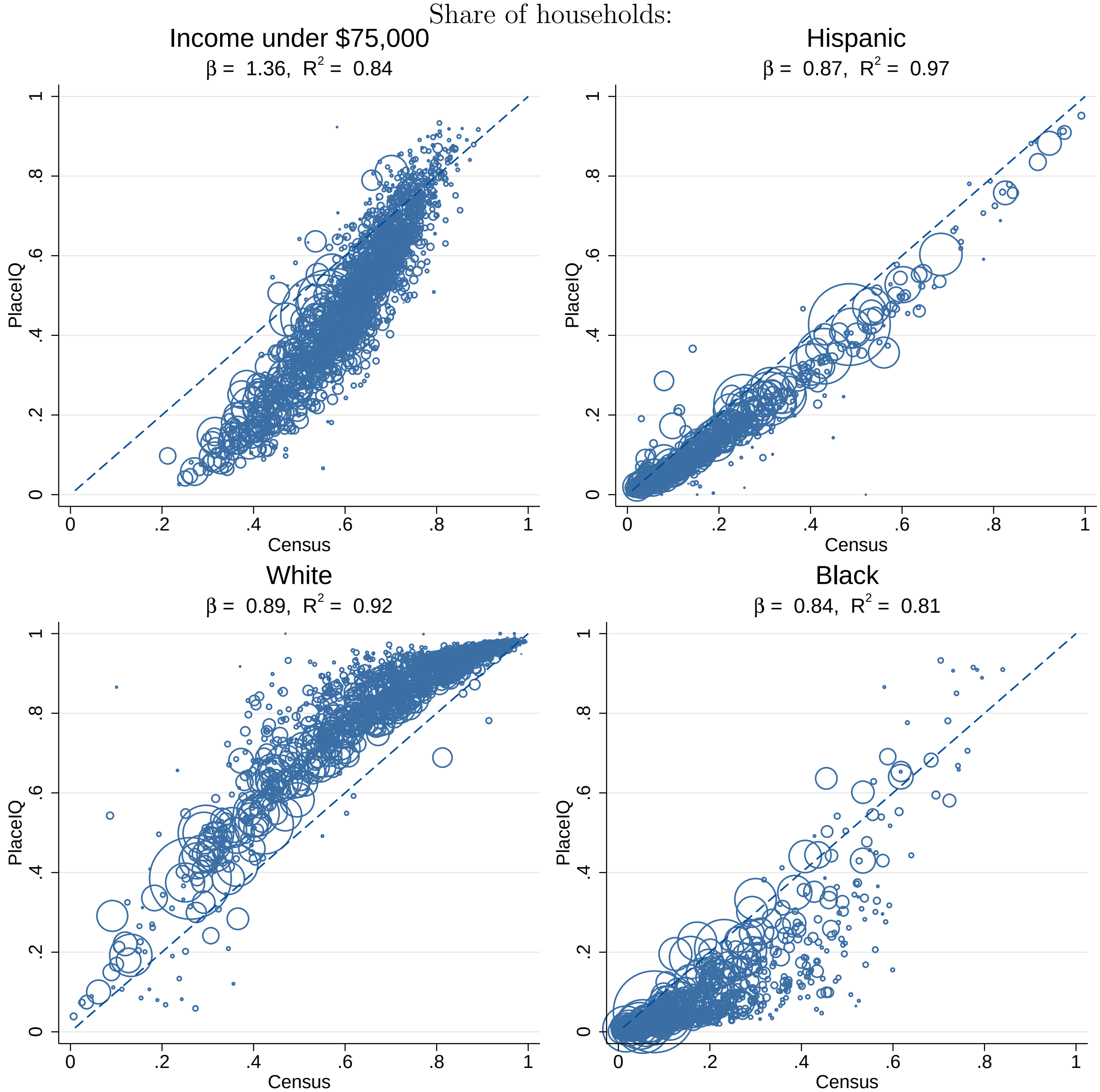
These differences also vary in intensity across counties. The top-left plot of Figure A.1 shows that the share of low-income devices in the building-level data is smaller than the share of low-income residents in the Census, except in counties with the largest low-income shares. This means that low-income households are under-represented in the building-level 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 Precisely PlaceIQ data. We note that the disparities between demographics in our smartphone sample and Census data are more pronounced for income than for race. Given these differences and the arbitrary nature of a dichotomous definition of "high-income", we avoid reporting income exposure shares in absolute terms. This decision is based on the notion that, although we can identify venues with higher shares of high-income individuals, we may systematically overestimate the income of individuals in these venues.

A.5 Building-level data reliability

We now seek to validate the accuracy of the demographic information in the building-level data. First, we compare the Precisely PlaceIQ building-level data to building-level incomes inferred from house prices in Cook (2023) and address-level race and ethnicity data from the North Carolina State Board of Elections (NCSBE) voter registration data. Then, we demon-

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

Share of households:



NOTES: These plots compare county-level demographic composition in the 2015-2019 American Community Survey and that of devices for which we have 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 Precisely PlaceIQ county shares on ACS shares, weighted by the ACS population.

strate that the Precisely PlaceIQ data reliably predicts differences in chain visit propensity between residents of neighboring houses (i.e., two households living in the same block group).

A.5.1 Comparison with imputed building-level income from Cook (2023)

We compare our building-level data to address-level income data from Cook (2023), who kindly shared block-level averages with us. The Cook (2023) income data is imputed from home parcel characteristics from CoreLogic – including market value, size, and location – and Census tables on block group income. To perform this comparison, we create a corresponding Census block-level dataset by averaging our building-level data over all buildings in a Census block and assuming a uniform distribution of income within each income bracket. The income data in Cook (2023) correlates with our own with a slope of 0.92 and an R-squared fit of 0.43, for Census blocks below the Precisely PlaceIQ topcode of $150,000.

Table A.2 focuses only on whether both datasets agree on above-median income classification. We regress a dummy for above-median income in our PlaceIQ Precisely data to a similar dummy in Cook's data. The R^2 of that regression, in column 1, is 0.21. The unmatched blocks rarely reflect wide discrepancies. Our above median-income classification agrees with that in Cook (2023) for 75% of census blocks, but allowing for a $10,000 buffer around median income raises the match rate to 91%. So even at the narrowest Census geography, our income data rarely disagrees with Cook's data by more than $10,000.

The second column of Table A.2 adds a block group fixed effect. The regression coefficient remains highly significant (t-statistic of 154), but the R-squared is much smaller at 1.1%. So while we are highly confident that the two datasets are related below the block-group level, the smallest geographic unit for which Census income data are available, we cannot tell from this comparison how noisy our building-level dataset is at small geographical scales. The data in Cook (2023) comes from home parcel data that is imperfectly correlated with income. Our building-level data only contains a fraction of the buildings in any given census block, so it is also measured with error.

A.5.2 Comparison with address-level race in North Carolina voter registration data

We now compare the Precisely PlaceIQ building-level race data to the address-level race data from the North Carolina State Board of Elections (NCSBE) voter registration data. In particular, we ask whether the Precisely PlaceIQ data can match racial composition at the address-level in the NCSBE data better than one could using the most detailed information available in the Census.

Table A.2: Agreement on block-level incomes in Precisely PlaceIQ and Cook (2023) data

	High-Income in Precisely PlaceIQ	
	(1)	(2)
High-Income in Cook (2023)	0.470	0.153
	(0.001)	(0.001)
Observations	2248831	2245465
Block Group FE	No	Yes
R-squared	0.212	0.510
Within R-squared	0.212	0.011

NOTES: This table reports the results of regressing a dummy for above-median income from our PlaceIQ Precisely data on a similar dummy from Cook (2023). Each observation is a Census block.

We first compute the share of buildings in which the race of residents reported in Precisely PlaceIQ and NCSBE are matching. To simplify this comparison, we compute this match rate using only monoracial buildings. So we identify buildings in the Precisely PlaceIQ data with all residents either White or Black, and we match them to addresses in the NCSBE data whose Census coordinates are within one meter of the Precisely PlaceIQ building. We limit attention to single-family homes in the NCSBE data and buildings with 10 or fewer devices in the Precisely PlaceIQ data. We then calculate the share of monoracial Black buildings in the Precisely PlaceIQ data in which only Black voters report residing in the NCSBE data (and, similarly, the share of monoracial White buildings in which only White voters report residing). This is a demanding exercise because we only record a match if all residents of a building match in both datasets, and the voter registration data is prone to errors.

The first column of Table A.3, 'Building-level' reports the share of monoracial buildings in the Precisely PlaceIQ data in which all NCSBE voters are of that same race. This match rate is 85% for White buildings and 61% for Black buildings. This implies, for instance, that in 39 percent of buildings, there is a least one non-Black registered voter in a building that Precisely-PlaceIQ reports as monoracial Black. These discrepancies could be due to errors in either the Precisely-PlaceIQ data or the voter registration data, errors in geocoding, variation in reporting of Hispanic status (we report Black as non-Hispanic Black) or of multiracial individuals, or the fact that both datasets were collected at the same time.

In other columns, we show that our ability to match NCSBE voters race would be worse if we used Census data instead of our building level data. For instance, if we simply allocated race to buildings in proportion to racial shares in the entire state of North Carolina, the share of Black building in which we correctly match voters racial composition would be only 0.18. If we used instead Census data at the block group level, the smallest Census geography

Table A.3: Match rate in racial composition of Precisely PlaceIQ and NCSBE Data

	Match rate when using data at:			
	Building level	Block level	Block Group level	State level
White	0.85	0.84	0.82	0.77
		(0.0003)	(0.0004)	(0.0009)
Black	0.61	0.58	0.49	0.18
		(0.0015)	(0.0026)	(0.0048)

Notes: The table shows the match rate between race of voters in the NCSBE data and our building-level data, and compares that match rate with predictions from Census tables at various geographies. The first row reports this match rate for monoracial White buildings and the second row for monoracial Black buildings. The first column shows the share of monoracial buildings in the Precisely PlaceIQ data in which a all voters in the NCSBE data report being in that same race. The second, third, and fourth columns report the mean and standard deviation (in parentheses) of these shares calculated for 50 simulated versions of the Precisely PlaceIQ data, in which building demographics are randomly re-assigned within Census block, Census block group and state, respectively. White and Black are defined as non-Hispanic White and non-Hispanic Black, respectively.

at which the interaction of race and income we use in the paper is available, the match rate rises to 0.49, still lower than 0.61 achieved with our building level data. Unlike income, race is available at the Census block level, but even if we allocate race to buildings using this most geographically detailed Census data—as in smartphone papers like Athey et al. (2021)—the match rate for Black buildings only rises to 0.58.

Overall, we match address-level racial composition in the North Carolina voters data better with our building level data than one could using Census data. For monoracial Black buildings, the match rate is (0.61-0.49)/0.61 = 20% better than what can be achieved using the Census block group data that we would otherwise have to use, and 5% better than using Census block data, the narrowest Census geography at which race (but not income) is available.

A.5.3 Internal validation from chain visit propensity

The Precisely PlaceIQ data reliably predicts differences in chain visit propensity between residents of neighboring houses (i.e., two households living in the same block group). For example, we find that residents of high-income buildings are more likely than their neighbors in low-income buildings to visit chains preferred by high-income individuals, such as Starbucks.

To show this, we compare visit patterns between buildings within the same block group. This exercise establishes that the building-level data is informative at a finer level than the most granular demographic data publicly available from the Census Bureau, which is at the block group level.

Our methodology is based on the idea that building-level demographic differences should 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). Consistent with this idea, research by Waldfogel (2008) and Klopack (2020) shows that race and income correlate with heterogeneity in preferences for different types of venues and chains.

These observable differences in behavior predicted using only building-level demographic data should, in turn, correspond to similar observable differences in behavior predicted using only demographic information from the Census. Therefore, we 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 ranking 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 only building-level demographic information. For each chain in each block group, we compute the ratio of the average number of visits by devices living in high-income buildings (at least 67% of residents earning more than $\$75,000$) to the average number of visits by devices living in all other buildings. We then take a weighted average of this ratio across block groups to obtain a ranking of the chains by popularity with high-income relative to non-high-income individuals.[45]

Second, we compute the within-tract chain popularity ranking using only Census block group demographic information. To do this, 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 all other block groups. Finally, we compute an analogous set of ratios for visits by White versus non-White devices..

Figure A.2 depicts the results of these comparisons for restaurants and convenience stores

[45] Let g_i be an indicator if device i belongs to demographic group g, and $\bar{g}_b$ be the share of devices in block group b belonging to demographic group g. We weight using the number of devices living in a block group (N_b) and a variance weight ($\sum_{i \in b}(g_i - \bar{g}_b)^2$):

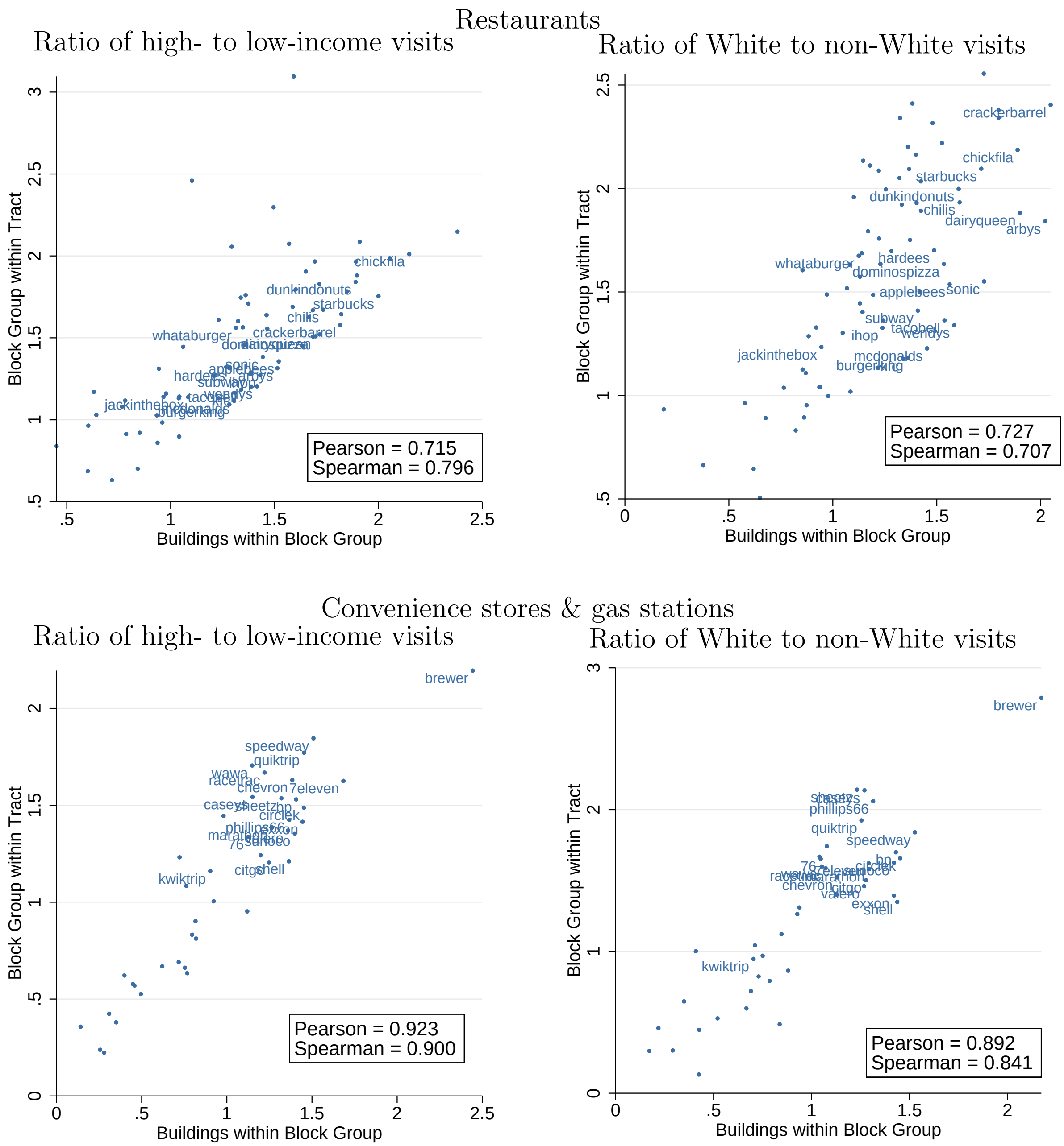
$$w_b = N_b \sum_{i \in b}(g_i - \bar{g}_b)^2.$$

Adding the second variance term produces a statistic that exactly matches the ranking of chains produced by the OLS estimate of γ_{cg} for each chain:

$$\log y_{ic} = \gamma_{cg} z_c g_i + \delta_{cb} z_c d_{b(i)} + \epsilon_{ibc},$$

where y_{ic} is the number of visits from device i to chain c, g_i is again an indicator if device i belongs to demographic group g, z_c is an indicator for chain c, and $d_{b(i)}$ is an indicator if a device i lives in block group b.

Figure A.2: Using building-level data to predict visit propensity within block groups



NOTES: The top-left panel depicts visits by high-income devices relative to visits by all other devices for restaurant chains. The vertical axis is this ratio for devices living in high-income relative to all other block groups within the same Census tract. The horizontal axis is this ratio for devices living in high-income relative to all other buildings within the same Census block group. The top-right panel depicts analogous ratios for White visits relative to non-White visits. The bottom two panels depict these comparisons using the "convenience store & gas stations" category rather than restaurants. The largest 20 chains by number of visits within each category are labeled on each plot. Data on block group income and race is from the 2015-2019 American Community Survey. High-income block groups are defined as having at least 67% of residents earning more than $75,000. White block groups are at least 67% White. Data on building-level income and race is described in this appendix. High-income buildings are defined as having at least 67% of residents earning more than $75,000. White buildings are at least 67% White.

& gas stations, the two business categories with the greatest number of chains. Each observation in the plot is a chain in those categories. Plots on the left-hand side show the relative propensity of high-income devices to visit a chain, and plots on the right-hand side 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 rank correlations between 0.7 and 0.9 for restaurant and convenience stores, for both income and race. For example, both building-level and Census-block-group-level demographic information suggest that high-income individuals make relatively more visits to Starbucks than Dunkin' Donuts.

Table A.4 summarizes information on the size of our venue sample for the five largest chains 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 Precisely PlaceIQ data. It also reports the total number of visits to each chain. The Precisely PlaceIQ basemap of venues is close to comprehensive, and contains upward 80 percent of all venues for most chains, except for gyms where the basemap is less comprehensive.[46] Bank and Gym chains receive fewer visitors than other categories, so preference estimates are noisier for these categories.

A.6 Google Places data

We use data from Google Places to assign characteristics to restaurant chain venues for the robustness exercise in Section 5. The Google Places venue data was collected by Akbar et al. (2023) in 2019.[47] That Google data is available in 98 of the 100 largest MSAs. The Google data identifies the name of the venue (e.g., McDonald's), as well as its exact geo-location, number of reviews, star rating, and price in four categories ($, $$, $$$, $$$$). We match a Google venue to a Precisely PlaceIQ restaurant chain venue if its location falls within the Precisely PlaceIQ polygon for that venue (enlarged by a factor 1.3 to account for uncertainty in the exact locations of establishments) and if the name of the chain matches. 53 percent of restaurant chain venues in Precisely PlaceIQ correspond to a Google establishment. In 86 percent of these cases Precisely PlaceIQ and Google establishment agree that the same chain is at the location. This match rate reflects in part the fact that the city boundaries in

[46] Three chains have more venues in the basemap than were open circa 2019 according to company records. For Rite Aid and Walmart, this reflects store closures rather than wrongly identified locations. We are less confident about why Safeway has more venues in our basemap, but it could reflect rebranding.

[47] That data comes from querying one of 109 keywords (like "restaurants") on Google Maps, at a large number of random locations within city boundaries defined by Akbar et al. (2023). The 109 keywords were designed to capture the universe of possible trip purpose.

Table A.4: Coverage of chain venues in Precisely PlaceIQ basemap

Category	Chain	Actual (#)	PIQ (#)	(%)	PIQ Visits (M)	Source
Bank	Bank Of America	4300	3124	73	0.68	
	Wells Fargo	5352	3035	57	0.66	
	Chase	4976	3225	65	0.57	
	PNC	2296	1687	73	0.51	
	Citizens Bank	1105	823	74	0.34	
Big Box	Walmart	3947	4429	112	393.25	
	Target	1868	1360	73	66.89	
	Costco	542	505	93	61.00	
	Sams Club	599	555	93	47.54	
	Tractor Supply Co	1844	1709	93	18.60	
Convenience	Shell	12846	12635	98	280.94	
	7-Eleven	9267	7712	83	204.16	
	Circle K	6339	6061	96	163.07	
	Exxon	10830	7633	70	153.78	
	Chevron	7892	7742	98	147.02	
Grocery	Kroger	2757	2247	82	103.73	
	Safeway	895	1315	147	35.56	
	Ahold Delhaize	1973	1524	77	31.50	
	Walmart Market	809	692	86	30.67	
	Publix	1239	855	69	28.20	
Gym	Planet Fitness	1934	809	42	0.88	
	LA Fitness	623	472	76	0.69	
	Orange Theory Fit.	1150	373	32	0.54	
	Anytime Fitness	2469	837	34	0.53	
	24 Hour Fitness	445	380	85	0.39	
Pharmacies	CVS Pharmacy	9895	8656	87	187.11	
	Walgreens	9168	8380	91	135.51	
	Rite Aid	2461	2649	108	23.40	
Restaurant	Subway	24154	21693	90	992.49	
	Starbucks	15041	10598	70	618.57	
	McDonalds	13846	13050	94	580.77	
	ChickFilA	2493	2030	81	224.12	
	Dunkin' Donuts	9630	7719	80	218.47	
	Burger King	7346	6789	92	153.43	
	Wendys	5852	5475	94	139.08	
	Taco Bell	6766	6743	100	138.03	
	Pizza Hut	7280	6085	84	129.12	
	Panda Express	2161	1630	75	101.41	

Notes: This table reports the number of venues within the five largest chains in all commercial venue categories: "Bank", "Big Box", "Convenience Stores & Gas Stations", "Grocery", "Gym", "Pharmacies", and ten largest chains for "Restaurant". The 'Actual (#)' column reports the number of U.S. chain locations as reported in various sources, which are hyperlinked in the last 'Source' column. The 'PIQ (#)' column reports the total number of venues in the Precisely PlaceIQ basemap, including those excluded from the estimation sample. The 'PIQ Visits (M)' column reports all visits to the chain between June 1, 2018 through December 31, 2019, by all devices in Precisely PlaceIQ, in millions.

Akbar et al. (2023) are smaller than those of MSAs, but also that the Google Places venue sample in Akbar et al. (2023) does not capture all venues available on Google Places.

A.7 National Household Travel Survey data

We use the 2017 NHTS (USDOT, 2017) to identify MSAs with large shares of consumption trips by car for the robustness exercises in Section 5. We match each NHTS household to our MSA boundary using their county of residence.[48] We only keep individuals aged 18 and over in the sample. We define consumption trips as including all trips with NHTS trip purpose code 11 (Buy goods: groceries, clothes, appliances, gas), 12 (Buy services: dry cleaners, banking, service a car, pet care), and 13 (Buy meals: go out for a meal, snack, carry-out). We only keep the 92 MSAs (out of the 100 largest MSAs) with at least 100 consumption trips in the NHTS. We define 'car' trips as all trips with NHTS transport mode code 3 (Car), 4 (SUV), 5 (Van) and 6 (Pick Up Trucks). We then use NHTS trip-level weight to compute the share of consumption trips by car. Among the MSAs with more than 100 consumption trips, there are 56 MSAs in which more than 90% of consumption trips are by car and 14 MSAs in which more than 95% of consumption trips are by car.

B Appendix tables and figures

B.1 Patterns of demographic exposure

[48]This geographic matching requires the confidential geo-coded version of the NHTS. We thank Gilles Duranton, who has access to it, for producing this list of MSAs for us.

Table B.1: Mean exposure to same-race co-patrons

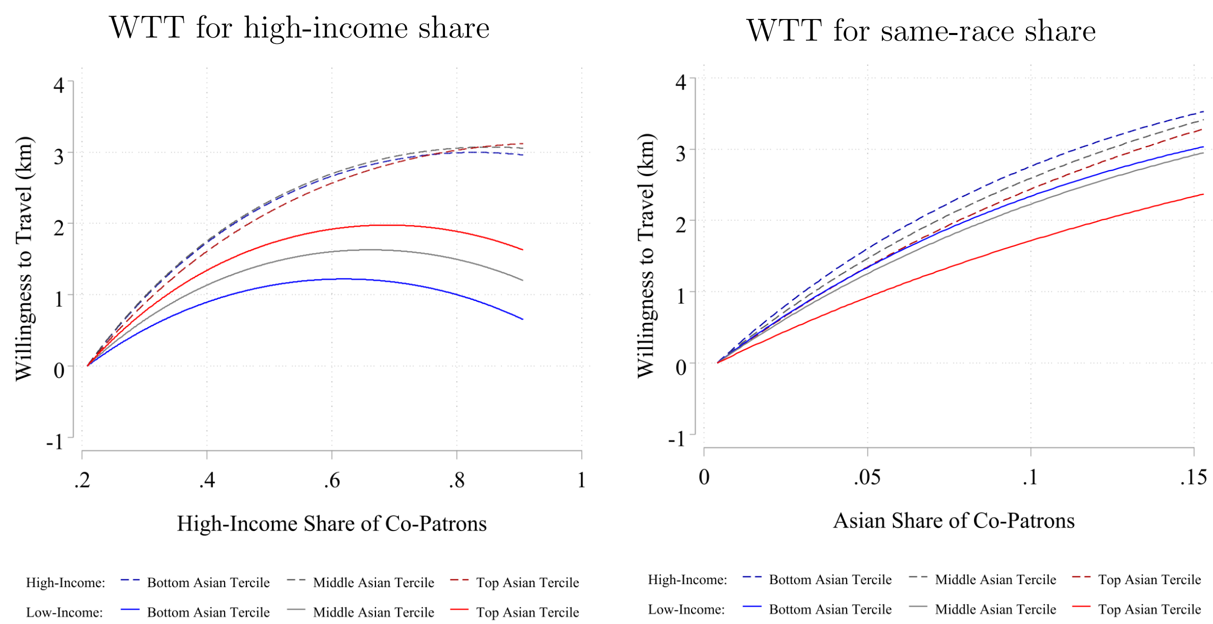
	Low Income				High Income			
	Asian (1)	Black (2)	Hispanic (3)	White (4)	Asian (5)	Black (6)	Hispanic (7)	White (8)
Estimation sample	0.08	0.27	0.25	0.08	0.08	0.19	0.13	0.11
All chain-restaurant visits	0.08	0.23	0.24	0.07	0.08	0.17	0.14	0.10
All chain-venue visits	0.09	0.25	0.24	0.07	0.09	0.18	0.14	0.09
All non-residence venue visits	0.15	0.30	0.32	0.02	0.15	0.21	0.19	0.05
All McDonald's restaurant visits	0.08	0.26	0.26	0.08	0.08	0.20	0.16	0.11
Census tracts	0.16	0.32	0.30	0.13	0.18	0.24	0.23	0.13

NOTES: This table is analogous to Table 1 but reports same-race exposure instead of high-income exposure. The table reports, for different visit samples, the same-race share of co-patrons that each demographic groups (eight columns) is exposed to, relative to a baseline in which all venues in that sample are visited with uniform probability. The first row shows those same-race shares for visits in the estimation sample. The second through fourth rows shows those shares for broader visit samples. The fifth row shows those shares only for visits to McDonald's restaurants. In the sixth row, those shares are computed as if each Census tract is a venue and individuals only visit the census tract that they live in.

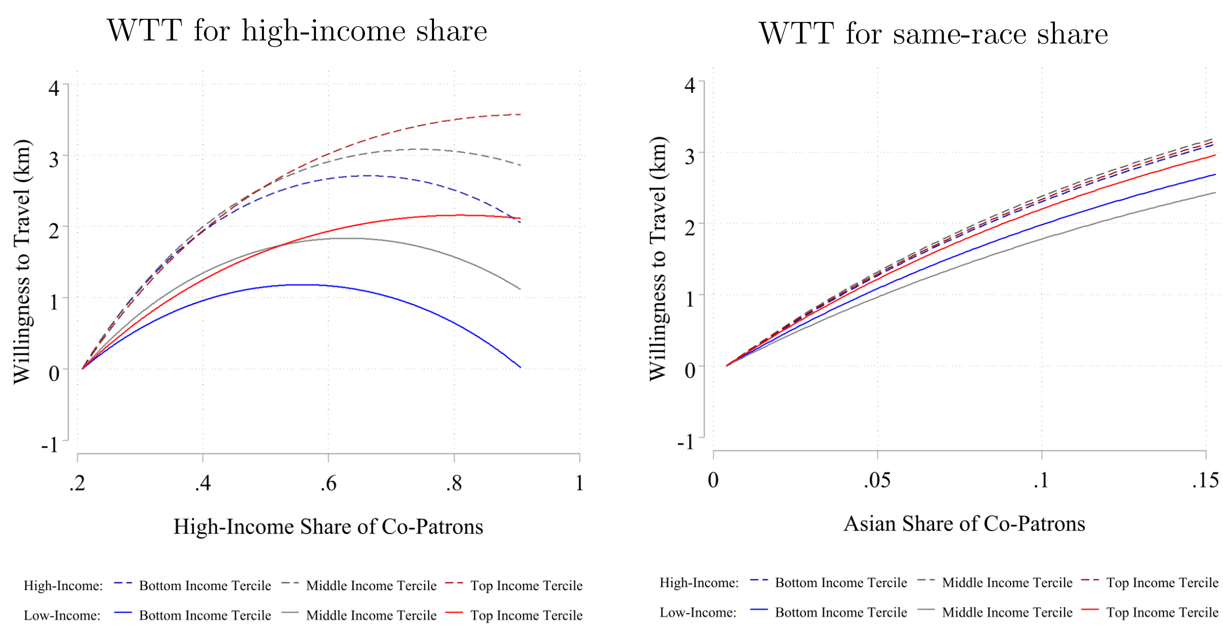
B.2 Determinants of demographic exposure

Figure B.1: Preference heterogeneity by residential neighborhoods

Asian visitors by residential tract's same-race tercile



Asian visitors by residential tract's high-income tercile

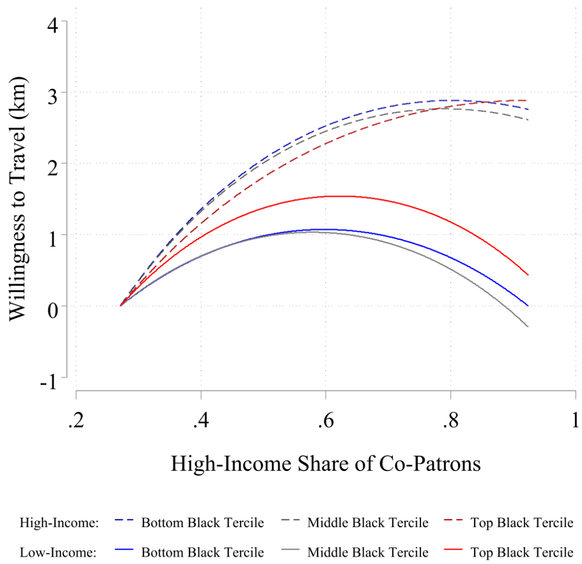


NOTES: This figure is analogous to Figure 5. except that it shows preference estimates for Asian, Black, and Hispanic visitors. Residential tract high-income terciles are defined using tract-level high-income share weighted by high-income tract population. Same for same-race terciles. The terciles for high-income residents are consistent across 8 demographic groups. *Continues onto next page.*

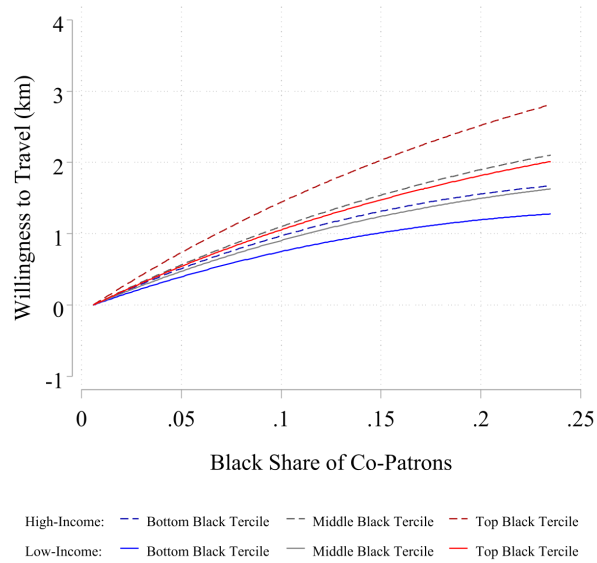
Preference heterogeneity by residential neighborhoods (continued)

Black visitors by residential tract's same-race tercile

WTT for high-income share

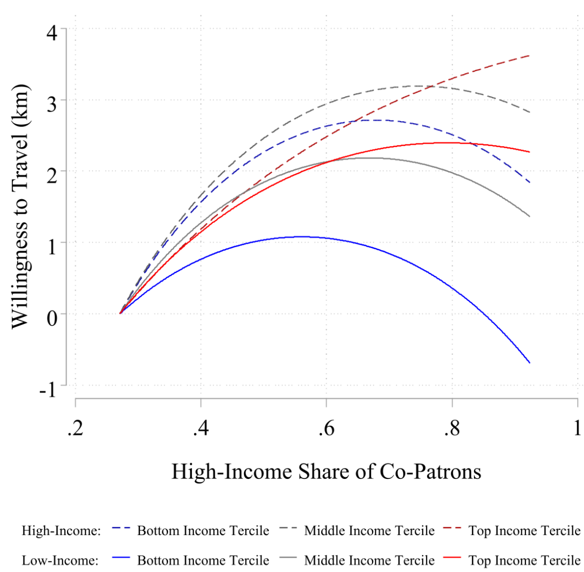


WTT for same-race share

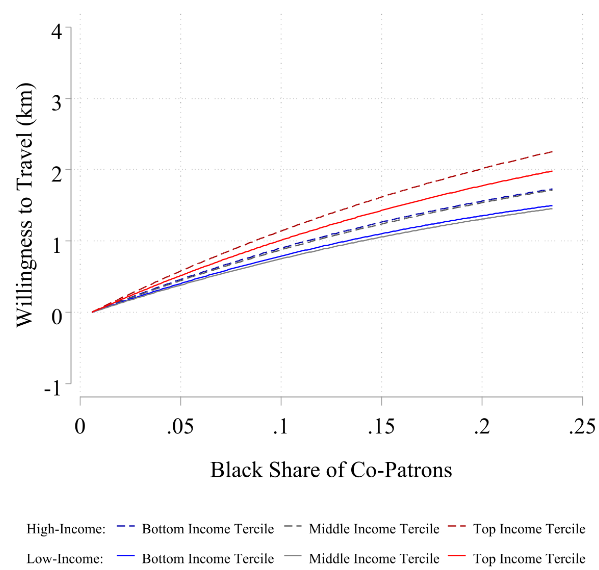


Black visitors by residential tract's high-income tercile

WTT for high-income share

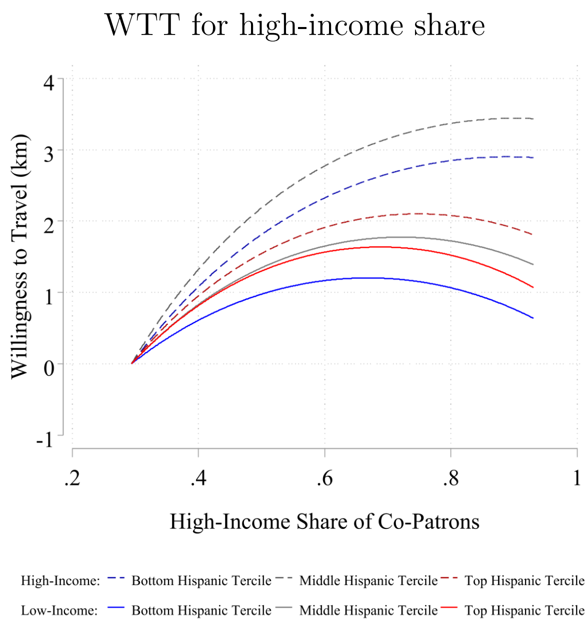


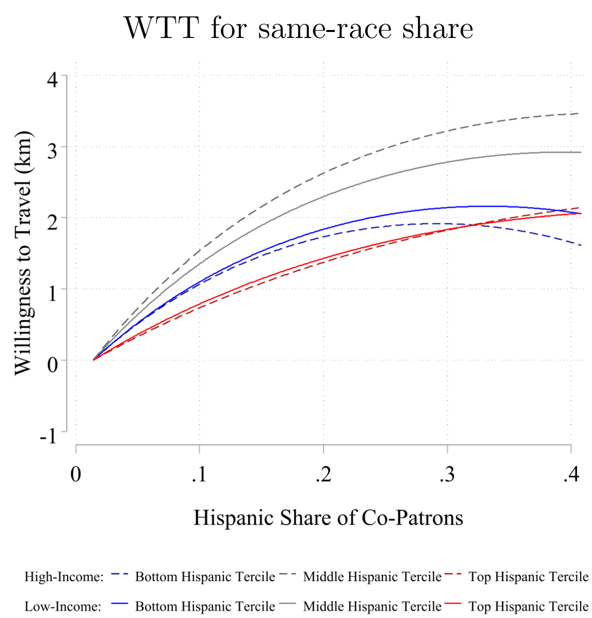
WTT for same-race share



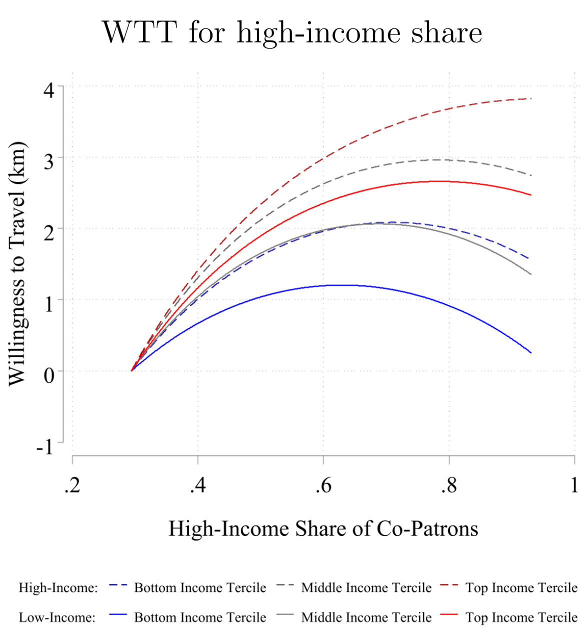
Preference heterogeneity by residential neighborhoods (continued)

Hispanic visitors by residential tract's same-race tercile





Hispanic visitors by residential tract's high-income tercile



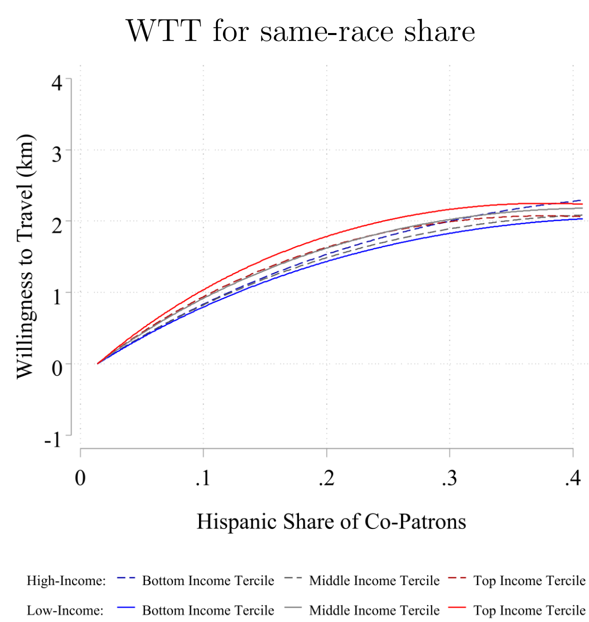
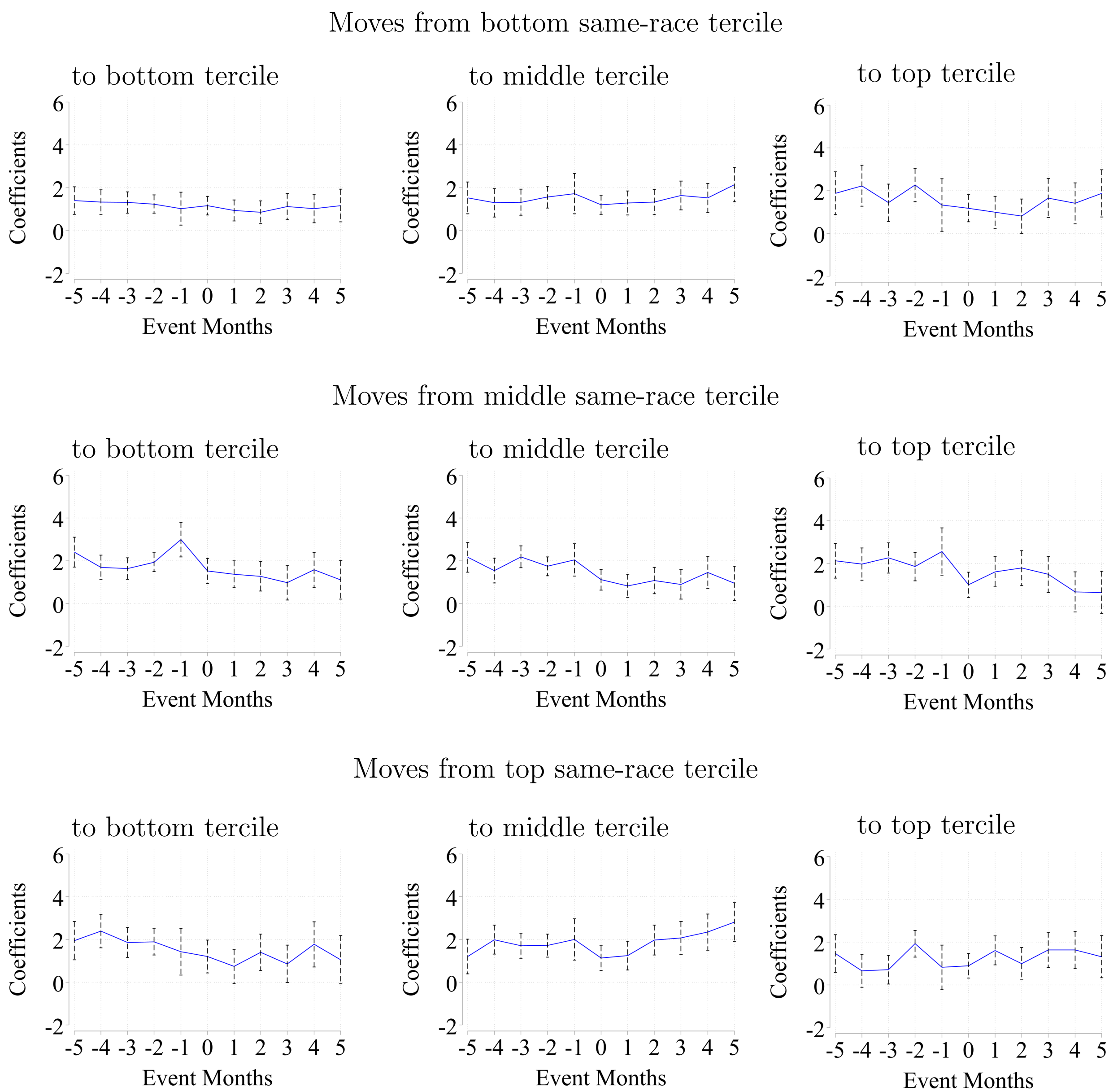
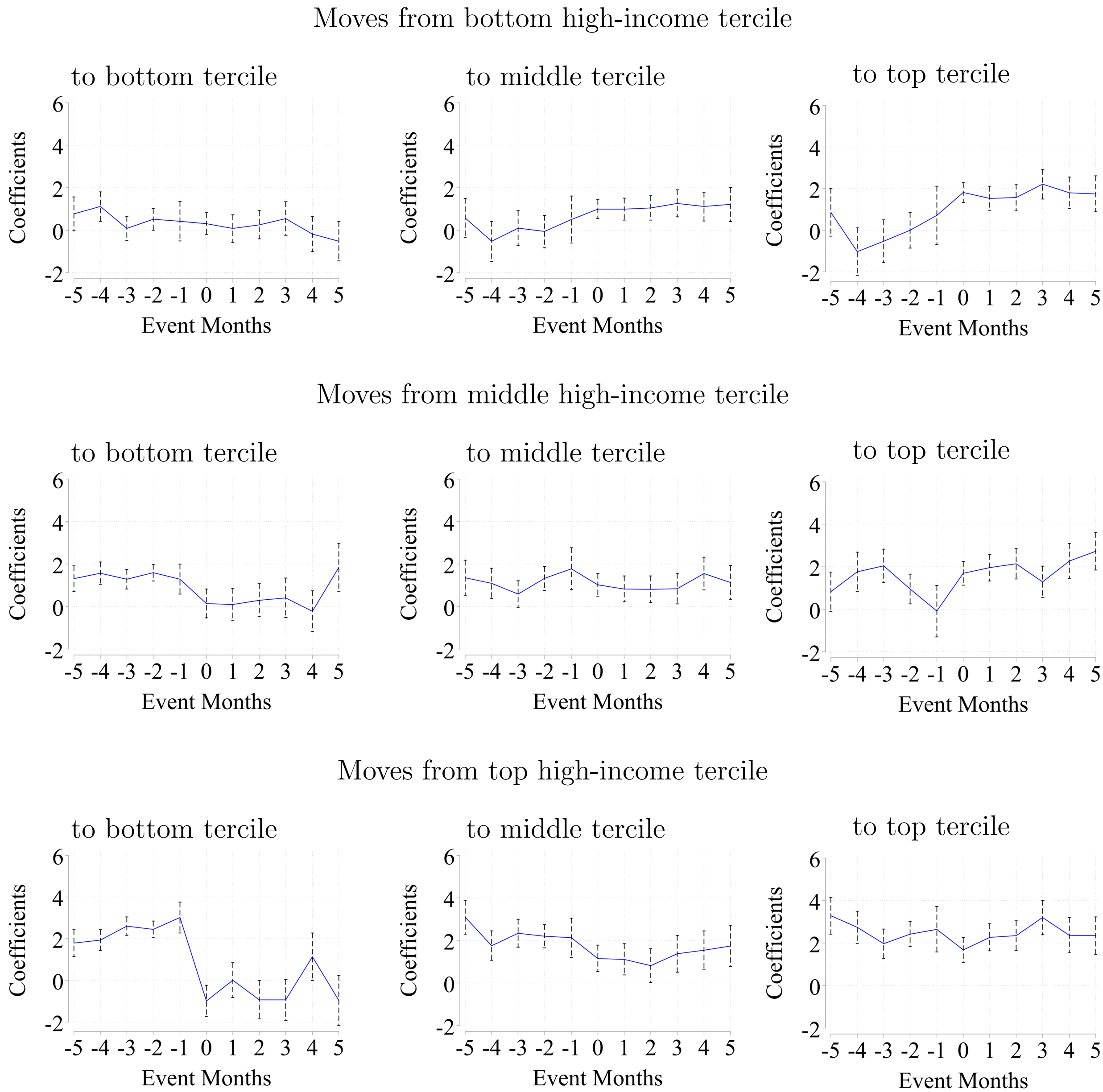


Figure B.2: Movers result across same-race terciles: High income preference



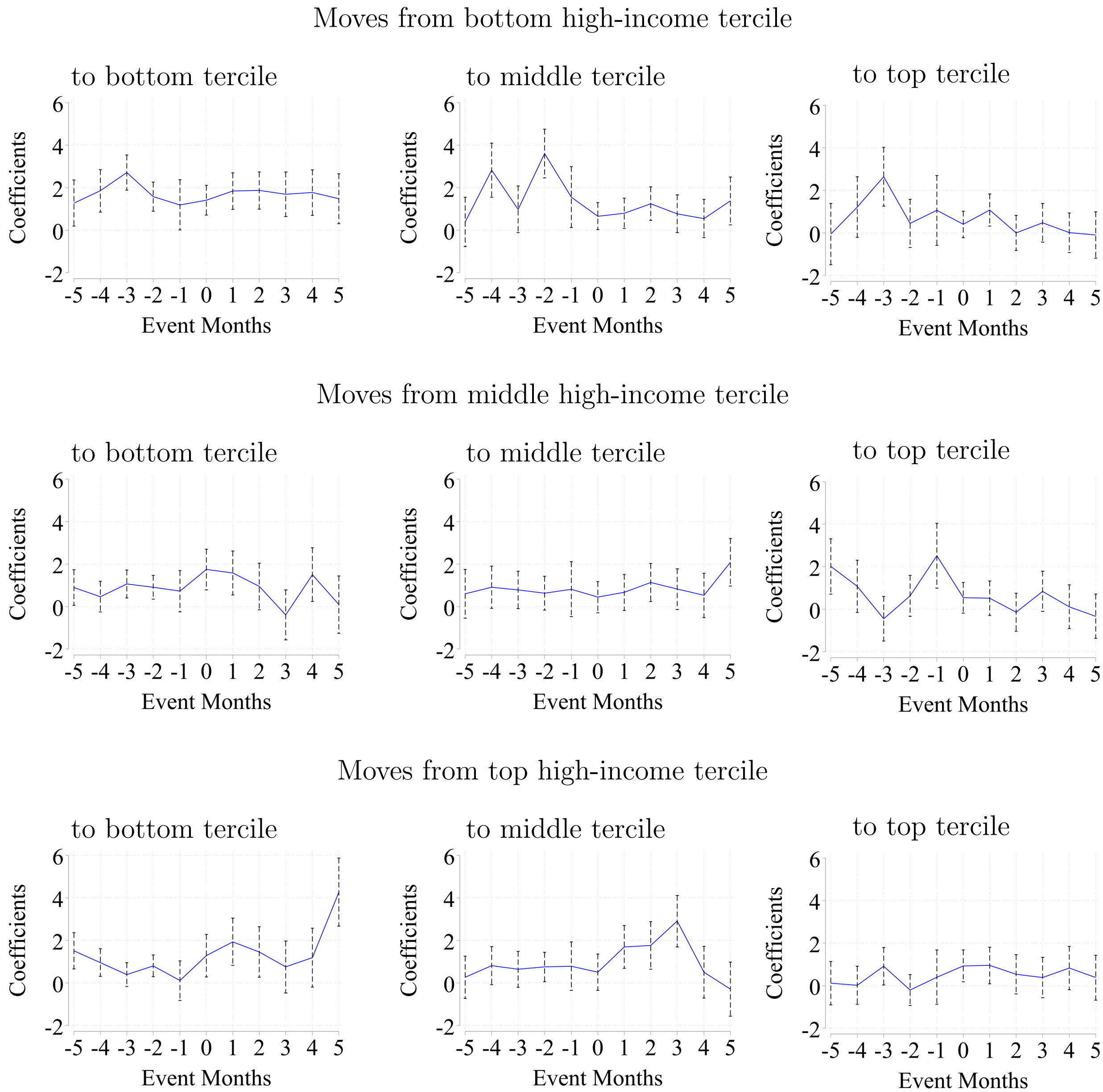
NOTES: This figure is analogous to Figure 6 except it shows estimates for high-income co-patrons $\beta_{yk}^{g,od}$ in equation (8).

Figure B.3: Movers result across high-income terciles: High income preference



NOTES: This figure replicates the exercise in Section 6.2.2 for moves between income terciles. It is analogous to Figure 6 except it showing estimates over high-income co-patrons $\beta_{yk}^{g,od}$ in equation (8) for cross-MSA moves between income terciles. Residential tract high-income terciles are defined using tract-level high-income share weighted by high-income tract population.

Figure B.4: Movers result across high-income terciles: Same-race preference



Notes: This figure replicates the exercise in Section 6.2.2 for moves between income terciles. It is analogous to Figure 6 except it shows estimates over same-race co-patrons $\beta_{rk}^{g,od}$ in equation (8) for cross-MSA moves between income terciles. Residential tract high-income terciles are defined using tract-level high-income share weighted by high-income tract population.

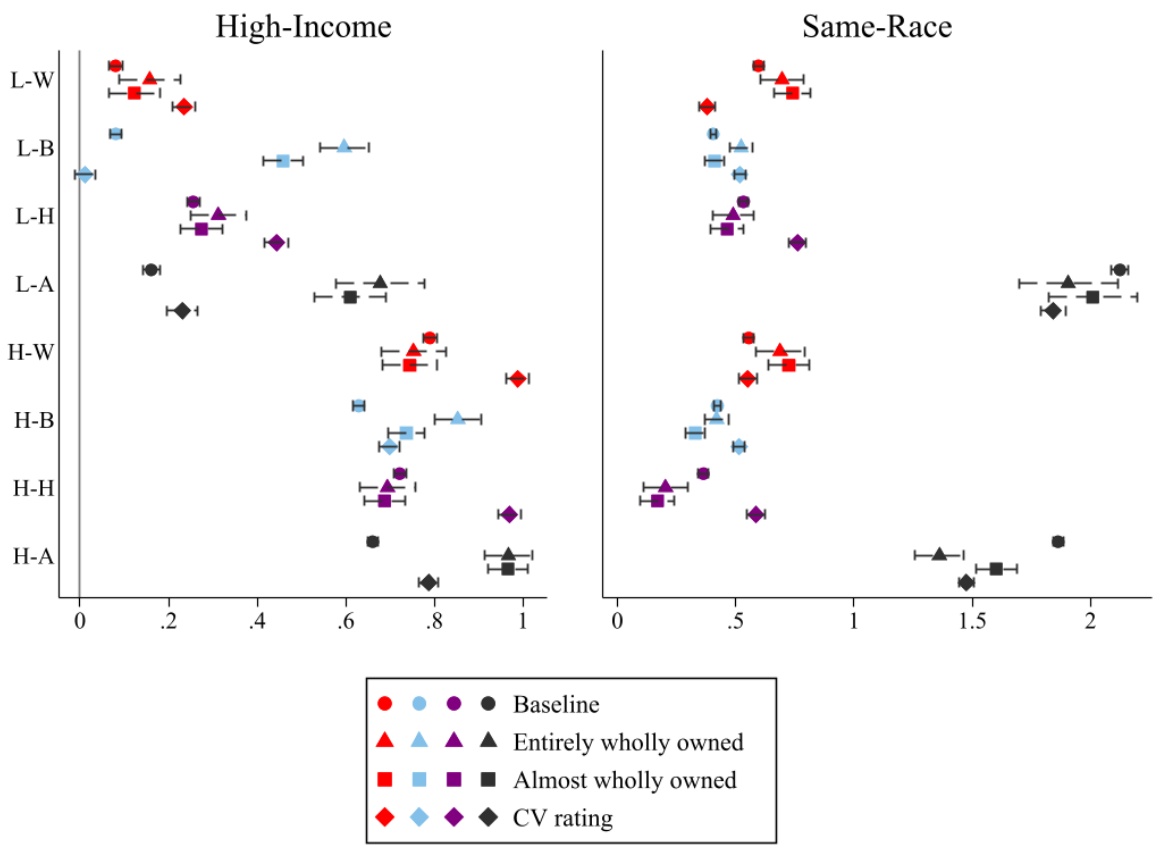
Table B.2: Average preferences of high-income White individuals before & after move across income terciles

	O1			O2			O3		
	D1	D2	D3	D1	D2	D3	D1	D2	D3
Pre-Move	0.58	0.12	0.01	1.42	1.23	1.10	2.36	2.30	2.61
	(0.16)	(0.21)	(0.26)	(0.12)	(0.17)	(0.21)	(0.13)	(0.17)	(0.19)
Post-Move	0.09	1.11	1.78	0.43	1.03	2.01	-0.45	1.29	2.37
	(0.15)	(0.13)	(0.14)	(0.18)	(0.14)	(0.15)	(0.20)	(0.17)	(0.16)

NOTES: This table is analogous to Table 4 except it shows estimates for moves between income terciles. The table reports pooled estimates of $\beta_{yk}^{g,od}$ from equation (8). For each origin-destination pair, coefficients are pooled for event-months prior to move ($t = -5$ to $t = -1$), and post-move ($t = 0$ to $t = 5$). The estimation samples contain home-venue-home visits to restaurants by high-income White individuals who move between MSAs, split by the high-income tercile of the origin residence (with O1 being the origin tercile with the lowest high-income share) and the high-income tercile of the destination residence (with D1 being the destination tercile with the lowest high-income share).

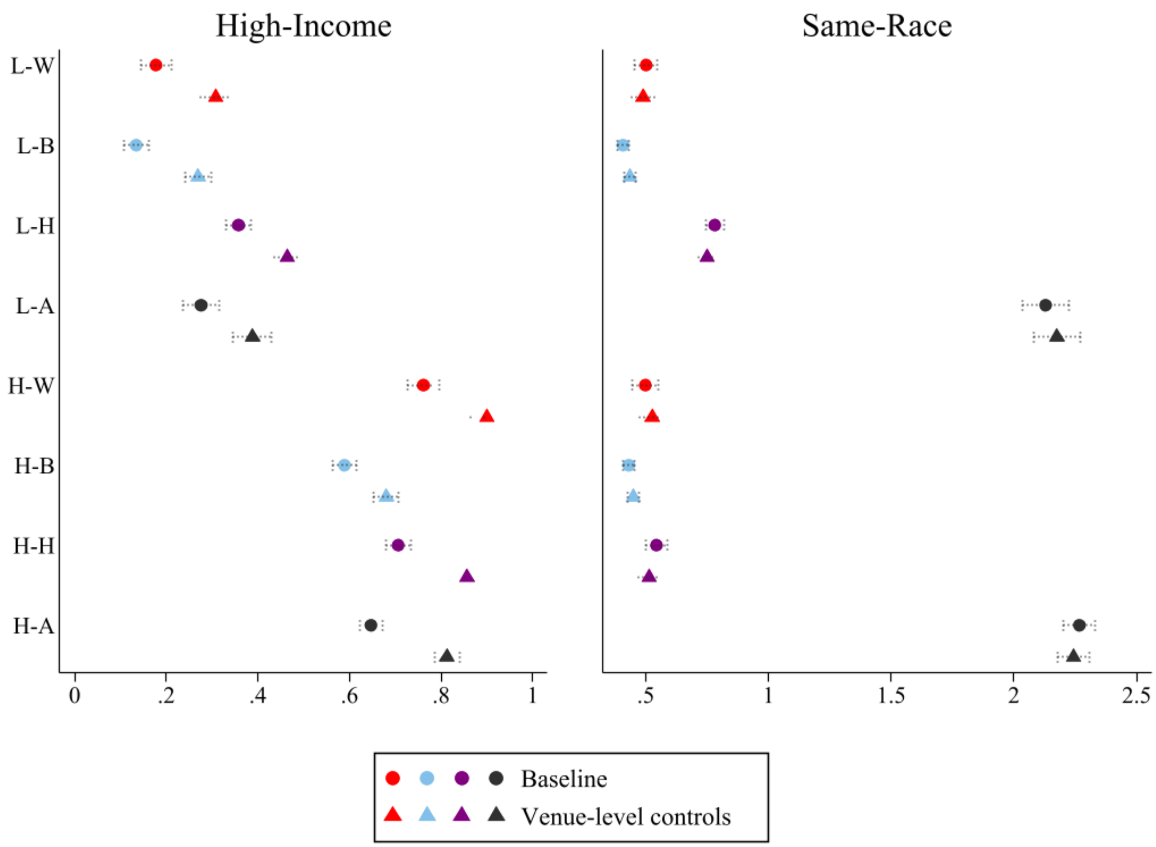
B.3 Robustness checks

Figure B.5: Robustness to standardized chains categories



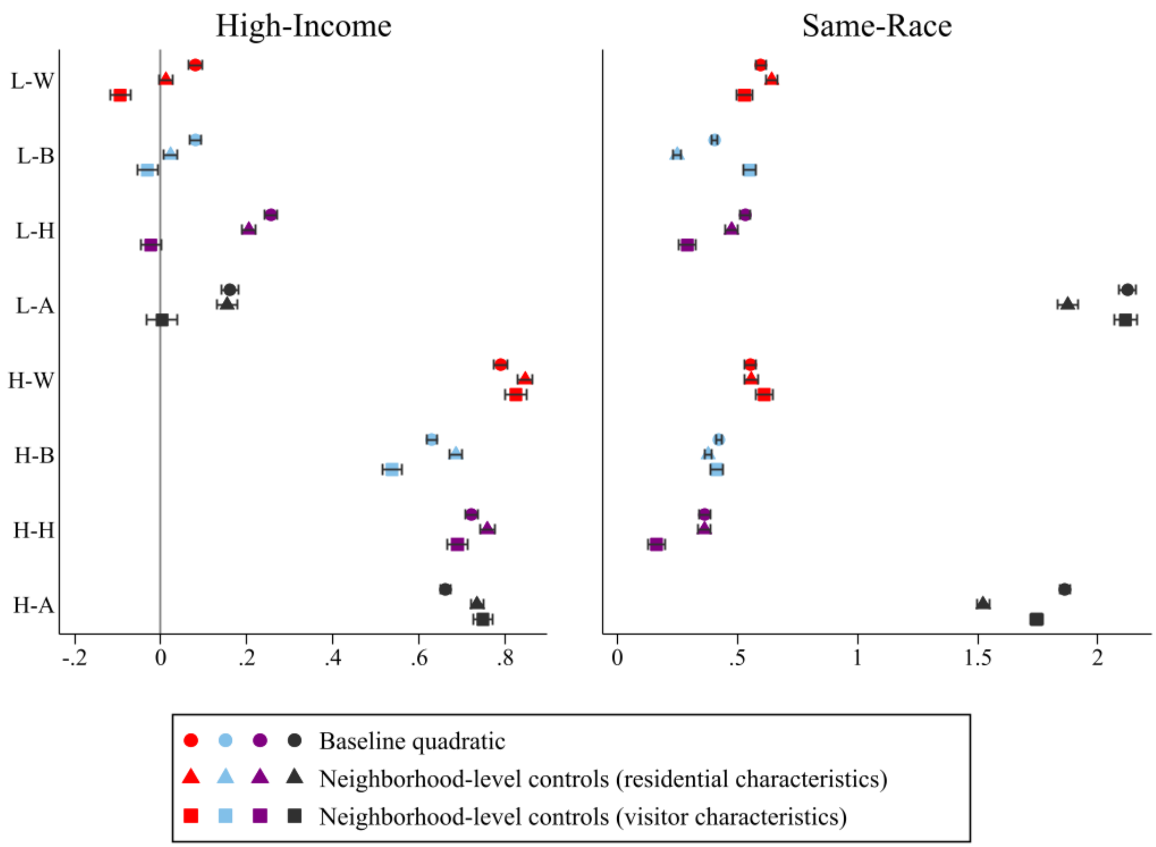
NOTES: These figures show results for restaurant chains restricted by chain standardization metrics. The preference estimates over co-patron composition use the linear shares specification from Equation (7) for each demographic group. Preferences are expressed in willingness to travel in kilometers relative to the average venue, $\Delta^g(s^{\text{samerace}}, s^{\text{highinc}})$, as defined in equation (3). The baseline sample includes preference estimates over all restaurant chains. "Entirely wholly owned" restricts to chains with 5% or fewer franchised venues; "Almost wholly owned" restricts to chains with 20% or fewer. "CV Rating" limits to the 25% of chains with the lowest variation in Google Places star rating.

Figure B.6: Robustness to adding venue-level controls



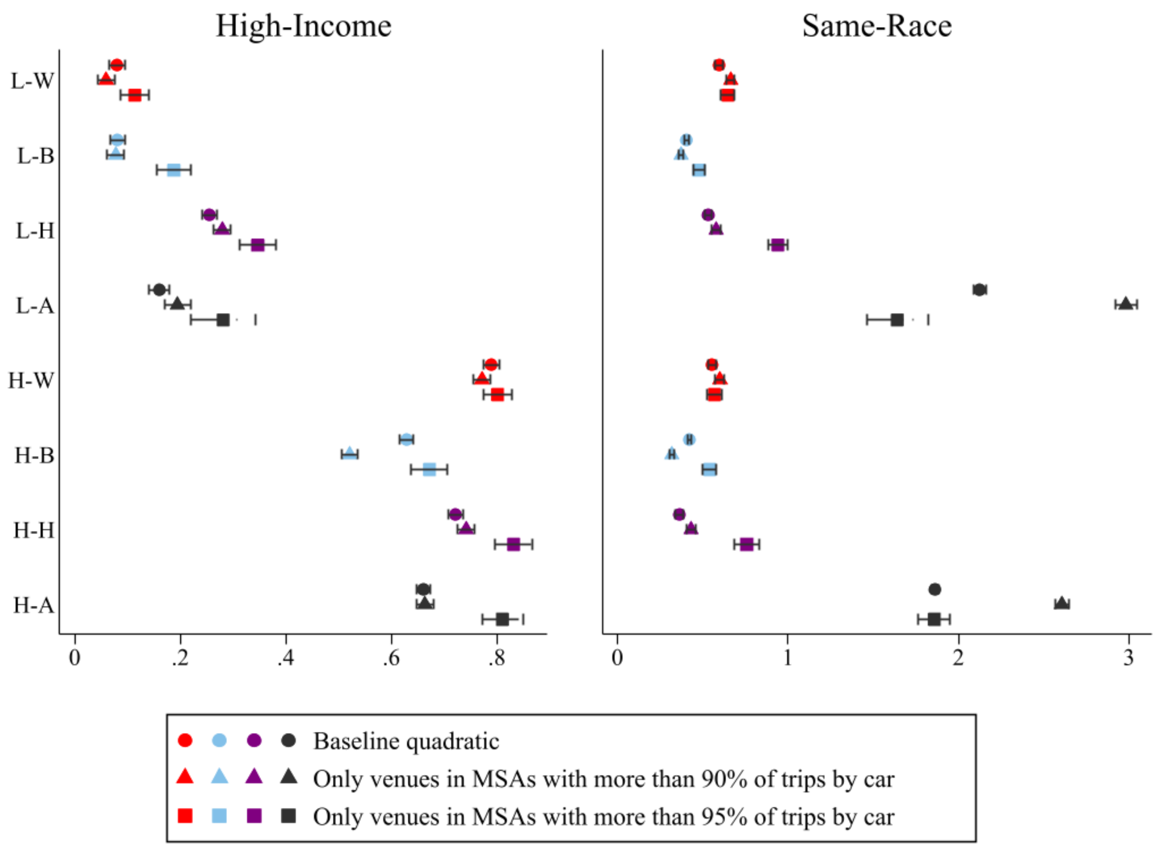
Notes: These figures are analogous to Figure B.5 except they add venue-level controls to the baseline estimation. Preference estimates over co-patron composition are reported from the linear shares specification from Equation (7). The venue-level controls are Google Places star rating, Google Places number of reviews, and venue square footage as reported by PlaceIQ Precisely.

Figure B.7: Robustness to adding neighborhood-level controls



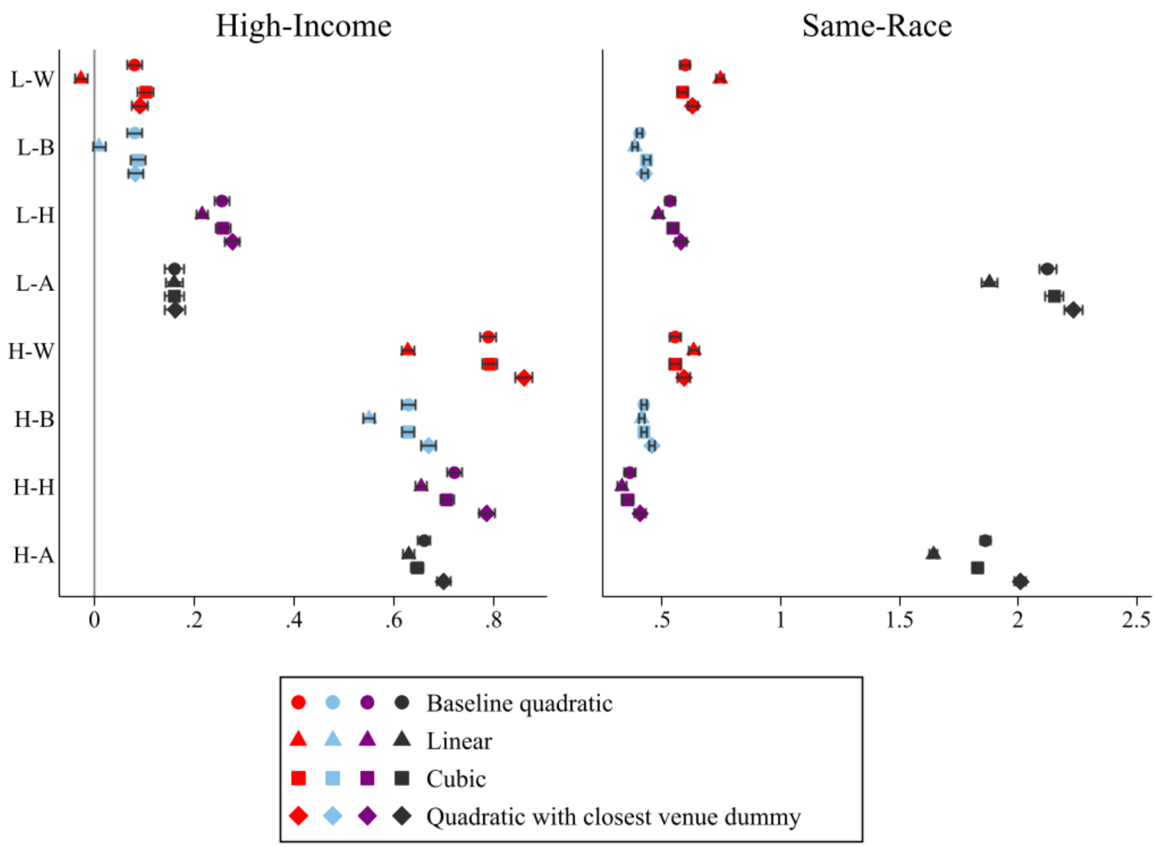
NOTES: These figures are analogous to Figure B.5 except they add neighborhood-level controls to the baseline estimation. The preference estimates over co-patron composition use the linear shares specification from Equation (7) for each demographic group. Residential characteristics include the shares of same-race residents and high-income residents in the census tract where each venue is located. Visitor characteristics include the same-race share of co-patrons and high-income share of co-patrons to all other commercial venues within the census tract where each venue is located. Commercial venues are identified by PlaceIQ Precisely.

Figure B.8: Robustness to transportation mode



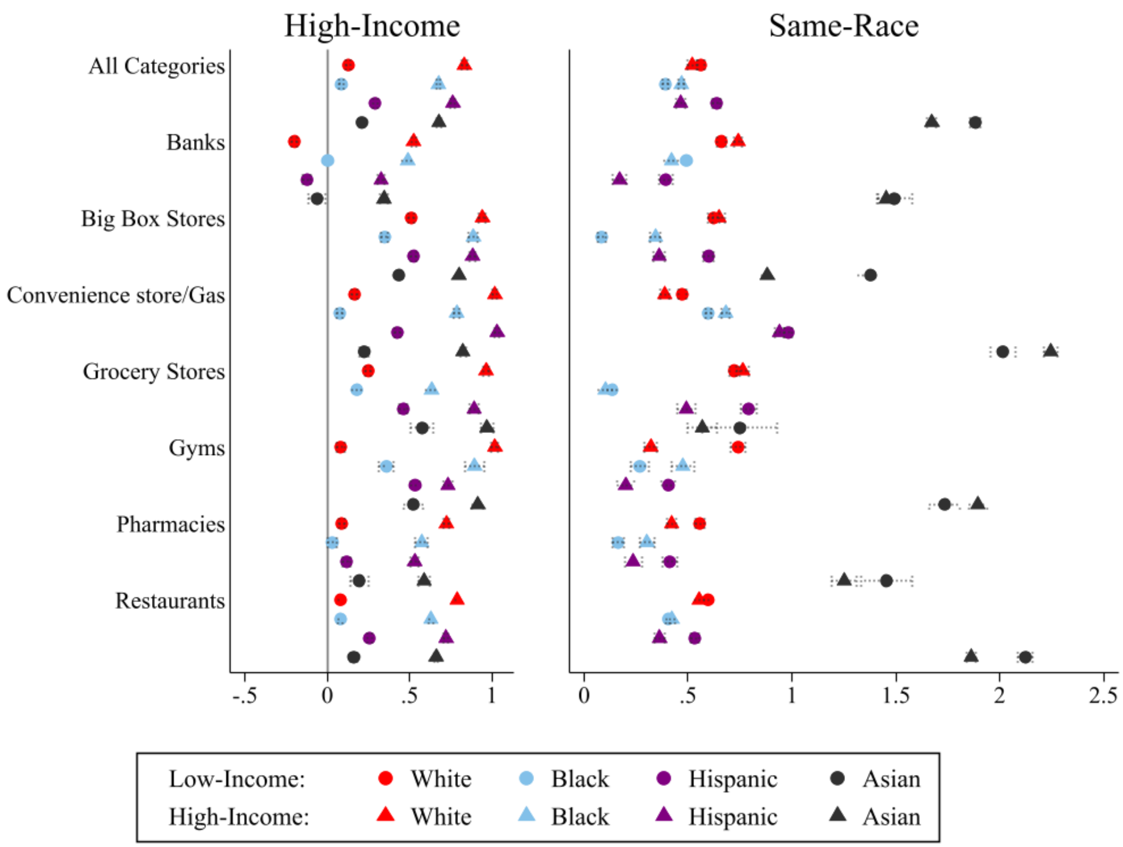
NOTES: These figures are analogous to Figure B.5 but restrict the set of MSAs in the baseline restaurant sample by car usage. Preference estimates over co-patron composition are reported from the linear shares specification from Equation (7). Using data from the 2017 National Household Travel Survey (NHTS), the sample of venues is limited to MSAs where at least 90% or 95% of trips to commercial venues are by car. For the 90% threshold, 79 MSAs remain, and for the 95% threshold, 23 MSAs remain.

Figure B.9: Robustness to distance specifications



NOTES: These figures are analogous to Figure B.5 but report preference coefficients when varying the specification on distance in the baseline restaurant sample. Preference estimates over co-patron composition are reported from the linear shares specification from Equation (7). The baseline quadratic specification is the same as reported in equation (2). The cubic specification adds a cubic log-distance term. The closest venue specification adds a dummy for the venue within each chain that is closest to a visitor's residence.

Figure B.10: Robustness to different categories



NOTES: These figures are analogous to Figure B.5 but report preference coefficients for various business categories. The preference estimates over co-patron composition use the linear shares specification from Equation (7). Each point represents the coefficient on the high-income share of co-patrons (left) or same-race share of co-patrons (right) for a specific chain category.