Choices or Constraints? Disentangling Preferences from Restrictions in Housing Markets*

Milena Almagro
Chicago Booth & NBER

Aradhya Sood
University of Toronto

November 5, 2026

Abstract

How much of observed household sorting is due to restrictions in choice sets, arising from *implicit* and *explicit* forms of discrimination, versus household preferences? What are the equilibrium and welfare consequences of restricting access for certain groups? To answer these questions, we build a two-sided housing matching model that combines admission rules with traditional discrete choice models, allowing for heterogeneous choice sets across demographic groups. To estimate the model, we construct a novel dataset that links households and real-estate developments to the historic street grid of the 1940 Minneapolis metro area. We jointly estimate heterogeneous choice sets and preference parameters finding that different demographic groups faced substantially different choice sets, shaped by both explicit and implicit restrictions. On average, non-White and White-ethnic households had access to 71% and 75% of neighborhoods, respectively, while non-minority households did not face any restrictions. In counterfactual simulations, removing these restrictions raises welfare for minority households, but rising prices dampen these gains. Household preferences over neighborhood demographic composition account for about three-quarters of the decline in non-White segregation that results from removing both restrictions and homophily preferences.

*Authors are listed in alphabetical order. All co-authors contributed equally. We thank the Mapping Prejudice team, and in particular Michael Corey, Kirsten Delegard, Kevin Ehrman-Solberg, and Penny Petersen, for sharing their covenants data and insights as well as the Urban Transition Historical GIS Project and John R. Logan for sharing their 1940 street grid for Minneapolis. We thank Dan Aaronson, Jason Abaluck, Nikhil Agarwal, Abi Adams-Prassl, Sara Bagagli, Pat Bayer, Giovanni Compiani, Rebecca Diamond, Mike Dinerstein, James Feigenbaum, Ed Glaeser, Rick Hornbeck, Samuel Myers Jr., Paulo Somaini, Chris Timmins, Dan Waldinger, Laura Weiwu, and seminar participants at the AREUEA National Meetings, AREUEA Virtual Seminar, CURE, Federal Reserve Banks of Chicago, Minneapolis, and Philadelphia, Junior Spatial Economics Conference at Sciences Po, NYU IO day, Midwest IO Conference, NBER SI (Urban), NBER IO of Housing Markets Conference, OSUS, Russell Sage Foundation Emerging Scholars Conference, Southern Economic Association, UofT/UBC/UQAM Canadian Real Estate Conference, UQAM, Mansueto Institute, Princeton, U Chicago Discrimination in the 21st century conference, LSE, Kings College, Tilburg, Rotterdam, Maastricht, Helsinki, Barcelona Summer Forum, SED, UCLA Spatial Conference, Urban Economics Association Conference, and Women in Empirical Micro for their helpful discussions and comments. This research is supported by the Russell Sage Foundation and Gates Foundation. Ben Couillard, Ben Fraser, Nanxi Fu, Jason Hong, Sarah Ji, Xiaofeng Liu, Jonathan Potter, Sriram Tolety, and Elias van Emmerick provided excellent research assistance. Emails: milena.almagro@chicagobooth.edu and aradhya.sood@rotman.utoronto.ca.

1. Introduction

A key empirical challenge in studying residential choice is that we usually only observe where households live, but not necessarily the set of neighborhoods available to them. If a household does not live in a particular neighborhood, is it because it prefers to live elsewhere, or because that neighborhood was not an available option? More broadly, the spatial concentration of a demographic group may reflect preferences for co-location, restrictions on residential choice sets, or both.

While most models of residential sorting typically assume that all neighborhoods are available to all households, a substantial literature in industrial organization and marketing studies how unobserved choice and consideration sets affect preference estimation (see surveys by Honka et al., 2019; Crawford et al., 2021). In housing markets, however, restrictions on choice sets can naturally arise through a variety of mechanisms. Before the 1968 Fair Housing Act, explicit discriminatory restrictions excluded certain demographic groups from particular neighborhoods (Aaronson et al., 2021; Sood and Ehrman-Solberg, 2026). Even today, landlords may discriminate when screening prospective tenants (Christensen and Timmins, 2023), while homebuyers may face screening and steering by real estate brokers and selective sellers (Christensen and Timmins, 2022; Gold et al., 2025). Financial, informational, or zoning regulations constraints can also restrict households' residential options (Bergman et al., 2024; Gallagher et al., 2024).

In this paper, we extend traditional discrete choice models of location choice to allow for restrictions on households' choice sets for two primary reasons. First, from a methodological perspective, ignoring choice set restrictions can bias estimates and welfare calculations (Christensen and Timmins, 2023). While the existence of this bias is a well established result (Honka et al., 2019; Crawford et al., 2021), we also derive conditions that determine the sign of this bias. Second, a significant yet unanswered question in urban and spatial economics is how making a portion of housing units off-limits to certain population subgroups affects prices, location choices, segregation, and overall welfare.The methodologies developed in this paper can be applied to other urban and spatial questions where it is important to distinguish between preferences and access constraints, and the

resulting equilibrium effects from choice constraints. Our historical setting is a practical application of these methodologies, investigating how residential sorting and equilibrium prices are shaped by both household preferences and restrictions on residential choice sets, arising from either *implicit* discrimination or *explicit*, policy-driven exclusion. This framework can be adapted to address contemporary issues of implicit discrimination and the financial or informational barriers that limit housing and neighborhood choices.

Specifically, we examine how much restrictions on neighborhood choice sets versus households' preferences shape the spatial sorting of households in the 1940 Minneapolis metro area. To do this, we construct a novel dataset that links households from the 1940 individual-count US census data and real-estate developments (neighborhoods) in the 1940 Minneapolis metro area to the street grid of that time. To our knowledge, this is the first dataset compiled at the development-household level in a historical context. Additionally, we link these data to information on racial covenants, an explicit mechanism of housing discrimination in our historical setting. These covenants, added to property deeds, explicitly prohibited certain racial and ethnic minorities from residing in covenanted homes and served as the primary *explicit* tool of housing discrimination in urban and suburban neighborhoods in the early to mid-20th century in the US.

The 1940 Minneapolis metro area offers two advantages for our analysis. First, the city expanded rapidly during this period. Real estate developers converted farmland into urban fringe and suburban developments with standardized housing and clearly defined boundaries. As a result, developers, rather than sellers of individual houses, often determined who could access a neighborhood. Second, racial covenants give us a direct measure of how restrictions on neighborhood access were set. Studies of discrimination rarely observe restrictions directly or have variables that proxy for them. Beyond covenants, many restrictions were also implicit in our setting. To capture them, our model includes variables that proxy for implicit restrictions, such as developers' admission practices in their other developments and financial constraints related to housing prices.

We show, theoretically and through Monte Carlo simulations, the importance of accounting for restrictions on choice sets in residential choice models. Using a simplified model, we show that wrongly assuming all households can access every option—a com-

mon assumption in location choice models—can lead to biased preference estimates if households face restricted choices in reality. Our theoretical findings indicate that the sign of the bias depends on how the accessible neighborhoods differ from the full set. Concretely, if the accessible neighborhoods have lower (higher) values of a characteristic than the full set, a model that assumes full access underestimates (overestimates) the taste for that characteristic. The misspecified model attributes the absence of choices in inaccessible neighborhoods to preferences rather than to access. For example, if financial constraints make expensive neighborhoods inaccessible, a model with full access overstates households' price sensitivity. Monte Carlo simulations show that the bias grows with the share of neighborhoods removed from households' choice sets and can even be large enough to flip the sign of the estimated preferences.

To account for neighborhood choice-set restrictions, we propose a two-sided matching model of the housing market in which choice sets differ across demographic groups, following the literature on selective admissions and latent choice sets (Goeree, 2008; Gandhi, 2026; Abaluck and Adams-Prassl, 2021; Agarwal and Somaini, 2025). Concretely, real-estate developers "accept" households into their developments according to a reduced-form admission rule. This rule depends on the household's race or ethnicity, so each demographic group faces a different probability that a given neighborhood enters its choice set. While our admission rule accounts for observable discrimination, such as that from racial covenants, it also includes variables that proxy for *implicit* restrictions, which are harder to measure. Concretely, it also includes the developer's propensity to admit minority households in its other developments, which captures implicit barriers, as well as housing prices, which capture financial constraints.

On the demand side, households from different demographic groups choose their most preferred neighborhood from the set of neighborhoods to which they are admitted. These choices depend on the household's preferences over rents, the neighborhood's demographic composition, and other neighborhood and location characteristics, which vary across demographic groups (Bayer et al., 2007).

We estimate the model by extending Goolsbee and Petrin (2004) and Bayer et al. (2007) to allow for the estimation of latent choice sets. To define our groups, we follow

the historical meaning of "minority," which in 1940 was not based on race alone (Hoyt, 1933; Brigham, 1923). Following Hoyt (1933), we classify households into three groups: *Non-White* households, who include all non-White households, regardless of birthplace; *White-ethnic* households, who are White immigrants from Southern and Eastern Europe, Asia, Africa, and Central and South America; and *Non-minority* households are U.S.-born White households and White immigrants from Northern and Western Europe. We validate this classification with reduced-form evidence on how racial covenants affected each group. Living in a fully covenanted development lowers the probability of being a non-White or an Eastern or Southern European household by 10 and 2.8 percentage points, respectively, and raises the probability of being a Western European household by 4.8 percentage points.

To estimate our model parameters, we face two identification challenges. First, we need to separate choice sets from preferences using the observed choices in the data. Under our model's assumptions, such as i.i.d. Type I extreme value demand shocks, functional form alone can, in principle, deliver identification. However, this argument attributes any departure from logit substitution patterns to the formation of choice sets. To rely less on our functional form assumptions, we follow Goeree (2008); Agarwal and Somaini (2025) and use an excluded shifter that affects the formation of choice sets but not household utility. Specifically, we assume that a developer's propensity to admit minority households in their other developments affects this development's admission rule but does not differentially enter utility across demographic groups, conditional on other covariates. The rationale is that developers may have preferences over who lives in their developments, beyond monetary incentives. Developers could then shape a development's demographic composition through whom they admit.

Second, we face two endogeneity concerns when estimating demand parameters, as both house rents and demographic composition within a neighborhood are equilibrium outcomes and can thus be functions of unobservable demand shocks. To address these endogeneity concerns, we construct two instruments: one based on steep land slopes that affect construction costs and, therefore, house rents (Saiz, 2010); and another using a shift-share migration instrument to instrument for the share of non-minority households

within a neighborhood (Card, 2001; Tabellini, 2020).

Our results show that access to neighborhoods differed sharply across demographic groups. Non-minority households were admitted everywhere, while White-ethnic and non-White households could access 75% and 71% of neighborhoods on average. Weighted by households, non-White access falls to 46%. Racial covenants reduced admission for both minority groups, and especially so for non-White, but restrictions extended beyond covenants. Developers who admitted more minority households in their other developments were more likely to admit them here as well. The pattern of access also differed between the two groups. Higher prices and larger developments reduced non-White admission but had no significant effect on White-ethnic admission.

On the demand side, the estimated price coefficients are negative for all three groups. Minority households are more price-sensitive, especially non-White households. We also find evidence that households preferred to live near others with similar demographic backgrounds. Accounting for restricted access matters for these conclusions. A model that uses racial covenants as the only source of neighborhood restriction, and a model that assumes every neighborhood is available, both yield misleading preference estimates.

We develop three equilibrium counterfactual scenarios. These remove racial covenants, all neighborhood choice-set restrictions, and both choice-set restrictions and preferences over neighborhood demographic composition. In all simulations, rents adjust to clear a housing market with inelastic supply.[1]

In the first counterfactual, we remove covenants. Covenants had small aggregate effects as segregation falls by 0.3% for non-White households and by 1.1% for White-ethnic households. By contrast, their local effects were larger. In neighborhoods with covenants, removing them raises the minority share by 13.3%. The reason for this local vs. aggregated contrast is that covenants were concentrated in a few areas, covering 9% of neighborhoods and 3% of lots in the metro area.

In the second scenario, we eliminate all choice-set restrictions and allow for full access. Relaxing these constraints increases housing demand, and rents rise by 9.4% on average.

[1]This is not a full spatial equilibrium counterfactual with endogenous demographic composition. Imposing such an equilibrium condition may amplify the role of preferences, as preferences over neighborhood demographics function like an agglomeration force.

Despite the rent increase, the number of minority households in the metro area rises by 22%, reflecting that the original restrictions disproportionately limited their mobility. Welfare rises by 13.5% and 12.4% of baseline rent for non-White households and White-ethnic households, respectively. Segregation, as measured by the adjusted dissimilarity index (Carrington and Troske, 1997), falls by 16% for non-White households and by 11% for White-ethnic households.

In the third scenario, we remove preferences over neighborhood demographic composition in addition to eliminating choice-set restrictions. Rents fall by 1.6% on average, but rent changes vary widely across neighborhoods. Rents rise in neighborhoods with high baseline minority shares and fall elsewhere. This pattern is driven by non-minority households. Without preferences over demographic composition, they move into neighborhoods they previously avoided. Because their demand responds little to prices and housing supply is fixed, rents in those neighborhoods rise sharply. The number of minority households in the metro area rises by 2.5%, less than the 22% increase under open access alone. Segregation falls by 66% for non-White households and by 97% for White-ethnic households. These findings suggest that preferences over neighborhood demographic composition play a larger role in sustaining segregation than choice-set restrictions. Removing restrictions accounts for about one-quarter of the decline in non-White segregation obtained when both mechanisms are removed. For White-ethnic households, it accounts for about one-tenth.

Literature review: Our paper contributes to the literature on screening and restricted choice sets in discrete choice models, neighborhood segregation, and the role of policy instruments in shaping the distribution of households across space in several ways.

First, non-price rationing and screening are widespread yet rarely modeled in choice models, particularly in urban and spatial economics (e.g., Waldinger, 2021; Cook et al., 2023). Our two-sided matching model of the housing market relates to recent advances in estimating preferences under unobserved heterogeneity in choice sets, developed primarily in industrial organization and marketing (Goeree, 2008; Conlon and Mortimer, 2013; Gaynor et al., 2016; Gandhi, 2026; Abaluck and Adams-Prassl, 2021; Agarwal and

Somaini, 2025).[2] Most studies of residential sorting instead assume that every household can choose every neighborhood. The closest papers to ours incorporate restricted choice sets into models of residential choice, measuring restrictions directly from a single mechanism: correspondence experiments on steering and discrimination (Christensen and Timmins, 2022, 2023) or observed mortgage approval rates (Ouazad and Rancière, 2019). We instead treat choice sets as latent, since we partially observe restrictions through racial covenants. We estimate them from observed choices and let several channels beyond covenants shape them: developers' admission practices, prices, and development size. Understanding these restrictions is essential to determine how much of observed sorting stems from constraints rather than from preferences over prices or neighborhood demographic composition.

In addition, the industrial organization literature on latent choice sets does not model how equilibrium characteristics respond to these restrictions (Conlon and Mortimer, 2013; Gandhi, 2026; Agarwal and Somaini, 2025). By contrast, our counterfactuals let equilibrium prices respond to changes in access. This allows us to measure the trade-off between gains from wider access and losses from price adjustments when evaluating welfare.

Second, we contribute to the residential choice literature by distinguishing between preferences and choice-set restrictions as drivers of household sorting. Understanding the causes of segregation matters for children's outcomes, upward mobility, and inequality in access to local public goods (Collins and Margo, 2000; Durlauf, 2004; Chetty et al., 2016; Trounstine, 2018; Chyn and Katz, 2021; Couture et al., 2024). Much of the existing research attributes racial and ethnic segregation to preferences for living near households of the same race or ethnicity, or to neighborhood design policies (Schelling, 1969; Fischel, 2001; **?**; Bayer et al., 2007; Card et al., 2008; Caetano and Maheshri, 2025; Bayer et al., 2025; Diamond, 2016). Historical evidence instead points to collective exclusion and discrimination (Cutler et al., 1999; Boustan, 2010; Shertzer and Walsh, 2019; Sood and Ehrman-Solberg, 2026; Akbar et al., 2025). Recent studies show that discrimination, as well as financial and informational barriers, also shapes neighborhood sorting today (Christensen and Timmins, 2022; Bergman et al., 2024). A key challenge in this literature

[2]See Honka et al. (2019) and Crawford et al. (2021) for surveys of this literature.

is to separate the effects of choice-set restrictions from those of household preferences, since both affect segregation. We quantify how much each shapes equilibrium outcomes, such as the spatial distribution of households and prices.[3]

Third, our paper contributes to the literature on how policy design shapes spatial sorting and housing prices. Much of this literature estimates localized effects with quasi-experimental and historical methods (Myers Jr, 1995; Shertzer et al., 2016; Rothstein, 2017; Trounstine, 2018; Fishback et al., 2023; Aaronson et al., 2021; Song, 2025; Hynsjö and Perdoni, 2022; Sood and Ehrman-Solberg, 2026). We instead study the entire metro area, which allows us to assess the aggregate effects of the factors behind sorting and price differences across space. Two exceptions are Bagagli (2025) and Weiwu (2025), who study segregation resulting from the US Interstate Highway System in a spatial equilibrium model. We advance this literature by incorporating choice-set restrictions that differ across demographic groups, using our new data and a two-sided matching model. We quantify the role of an *explicit* discrimination instrument, racial covenants, as well as that of *implicit* restrictions, which are harder to measure and which the model captures through group-specific admission rules. Separating the two matters for policy evaluation. In our setting, racial covenants account for under 2% of the effect of restrictions on non-White segregation, while restrictions as a whole account for about one-quarter of the decline obtained when preferences over neighborhood composition are also removed.

2. Background, Data, and Empirical Evidence

Our focus on the 1940 Minneapolis metro area is based on two key reasons. This period was characterized by rapid urban expansion, during which developers, rather than individual sellers, often admitted households into neighborhoods. This allows us to construct housing restrictions for entire neighborhoods that will later shape choice sets. Additionally, 1940 marked the peak use of racial covenants, providing a measure of *explicit* restrictions, in addition to the implicit restrictions on households' choices.

[3]Related work by Li (2023) examines choice-set restrictions assuming no selection in their formation. By contrast, we model and estimate admission rules in which a household's demographics and neighborhood characteristics determine its probability of being admitted to a neighborhood.

2.1 Historical Background on Housing Restrictions

In 1940, various subgroups within populations faced both implicit and explicit restrictions in the housing market, which restricted their choices of neighborhoods. Implicit restrictions could take the form of uncoordinated actions by developers, real estate agents, and house sellers, which likely steered households away from certain neighborhoods (Massey, 1990; Rothstein, 2017; Christensen and Timmins, 2022). There were also instances of collective group action, such as threats of violence (Albright et al., 2021), that made some developments unwelcoming to racial and ethnic minorities during that time.

In addition to the implicit restrictions, the first half of the 20th century in the US also saw explicit restrictions in housing markets through racial covenants, which were used to limit the presence of minority households and shape neighborhood demographics. These private agreements, between real estate developers, homeowners, and neighbors, prohibited racial, ethnic, and religious minorities from living in certain areas across urban and suburban America (see Appendix Figure F.1). In US metro areas that were largely urbanizing during the first half of the twentieth century, including the Minneapolis metro area, developers added almost all racial covenants at the subdivision stage of new developments and widely advertised the presence of racial covenants in their developments (see Appendix Figure F.2 for an example of such an advertisement). The enforcement of these covenants depended on "injured parties" (usually neighbors) suing those who violated them, along with real estate agents and developers discouraging minorities from moving into covenanted neighborhoods.[4]

Four key historical events established racial covenants as a primary tool for neighborhood design in the US during the early 20th century making them increasingly popular over time. First, the 1917 *Buchanan v. Warley* US Supreme Court ruling prohibited racially segregated neighborhoods through zoning, making private racial covenants more appealing. Second, the National Association of Real Estate Boards amended its charter in 1924 to promote the use of these covenants. Third, the Supreme Court reaffirmed their legal enforceability in the 1926 *Corrigan v. Buckley* case, leading to widespread adoption. Finally, the 1938 Federal Housing Administration's underwriting manual favored mort-

[4]See Sood and Ehrman-Solberg (2026) for a list of court cases related to racial covenants in the US.

gages linked with racial covenants. However, the 1948 *Shelley v. Kraemer* ruling declared these covenants unenforceable, effectively ending the use of racial covenants to design segregated neighborhoods.

2.2 Data

We combine data on historic real estate developments, developers, the historical street grid, racial covenants, the 1940 US individual count decennial census, and housing characteristics with other geographic data, which allows us to map households and racial covenants to the real estate developments and geography of the 1940 Minneapolis metro area. More detailed information about how we construct our dataset can be found in Appendix A.

Data on Historic Real-Estate Developments: Data on historic developments that have survived is available from the Hennepin County Assessor's Office in Minnesota, where the Minneapolis metro area is located. However, relying solely on available data for existing developments ignores those that were demolished or altered between 1940 and the present, leaving some 1940 households in the sample unmatched to any development. To address this issue, we gather information from historical city atlases for the municipalities of Minneapolis (1940), St. Louis Park (1939), Edina (1940), Richfield (1941), Hopkins (1954), and Robbinsdale (1941). These atlases contain to-scale maps of each development within these municipalities (see Figure F.3 for an example page). We then digitize and geocode the historical atlases to create a geo-located map of each missing development from the 1940s. There are a total of 3,158 developments in the Minneapolis metro area in 1940. For each development, we also digitize the number of lots as well as the average lot size.

Data on Developers and Subdivision Years: We digitize development plat maps (see Appendix Figure F.4 for an example), which record the developer or development firm associated with each development and the year in which the development was first subdivided. We standardize developer and firm names across records to link developments associated with the same developer, including cases in which an individual appears under a firm name. See Appendix A.3 for details on liking developers to developments.

Historic Street Grids and Enumeration District Boundaries: Linking households, developments, and covenants to the present-day street grid results in fewer than 41% successful

matches. This low match rate is largely due to changes in the historic street grid over the past 80 years, driven by highway construction and new residential and non-residential developments. To overcome this problem, we digitize and geocode the 1940 street grids for Minneapolis and five neighboring suburban municipalities—St. Louis Park, Edina, Richfield, Hopkins, and Robbinsdale—using city atlases and the Sanborn Insurance maps from around 1940. Together, these six municipalities form a contiguous part of the Minneapolis metro area and represent approximately 93% of Hennepin County's 1940 population. Additionally, we geocode and digitize the 1910-1940 enumeration district (ED), a larger than census tract geographic area assigned to a single census enumerator, boundaries for Hennepin County.

Data on Racial Covenants: The data on racial covenants comes from the Mapping Prejudice Project, which provides a complete census of all covenanted sales deeds and development maps in Hennepin County. We combine data from sales deeds and development maps to determine the share of real estate development that was covenanted. In addition to the location of 24,642 racial covenants, Mapping Prejudice also provides information on the date of covenant execution.[5]

Individual-Count US Census Data: We use the individual count data from the 1940 US decennial census for the six municipalities in the Minneapolis metro area in our sample. The data include information on an individual's address and enumeration district. It also includes a list of demographic variables such as race, country of birth (to determine ethnicity), occupation, marital and ownership status, number of household members, head of household. Since we conduct our analysis at the household-development level, we assume that the race and ethnicity of the head of household represents the entire household. We believe this assumption is reasonable, given that endogamy (marriage within the same race or ethnicity) was common in 1940. We also confirm this trend using the individual count data.[6] There are 164,674 households in the six Minneapolis metro area municipalities in 1940.

[5]Mapping Prejudice Project analyzed roughly three million pages of sales deeds and development maps in Hennepin County, covering the period 1900-1960. Five people double-checked the existence and date of the covenant execution for each deed and map, limiting the scope of error in data collection.

[6]99.8% of households in our sample marry within their own race or ethnicity.

From these data, we also obtain the household's tenure status—owner-occupied, renter-occupied, or group quarters. We observe monthly rental prices and housing values for renter-occupied and owner-occupied households, respectively; however, no housing price data are available for households in group quarters. To enable a comparison between rents and housing values, we impute the monthly rental value of owner-occupied housing using the methodology following Katz (2017). The imputed rental value of owner-occupied housing, combined with rental price data, yields the mean monthly rental value for 98.2% of the developments in our sample.[7]

Data on Housing Characteristics: We use lot size to measure housing quality and build year to assess the newness of a development. For extant houses, we obtain lot sizes from the county's assessor's office. For historical developments or destroyed houses, we refer to scaled city atlases, which provide detailed information on lot shape and size (see Figure F.3).[8] Build years are also sourced from the county assessor for the extant houses to date.[9]

Other Geographic Data: We digitize an 1899 map of Hennepin County to highlight wetlands and semi-submerged areas, enabling us to assess the underlying land quality in the region's lake geography. Additionally, we obtain slope data from the Natural Resources Conservation Service's comprehensive survey of Hennepin County conducted in 1953. We also digitize a 1933 map highlighting the locations of heavy and light industries zoned areas, as well as commercial zoned areas in the city of Minneapolis. Last, we digitize the 1913 streetcar map of the Minneapolis metro area. See Appendix Table A.3 for summary statistics at the household, development, and enumeration district level.

2.3 Construction of the Household–Neighborhood Dataset

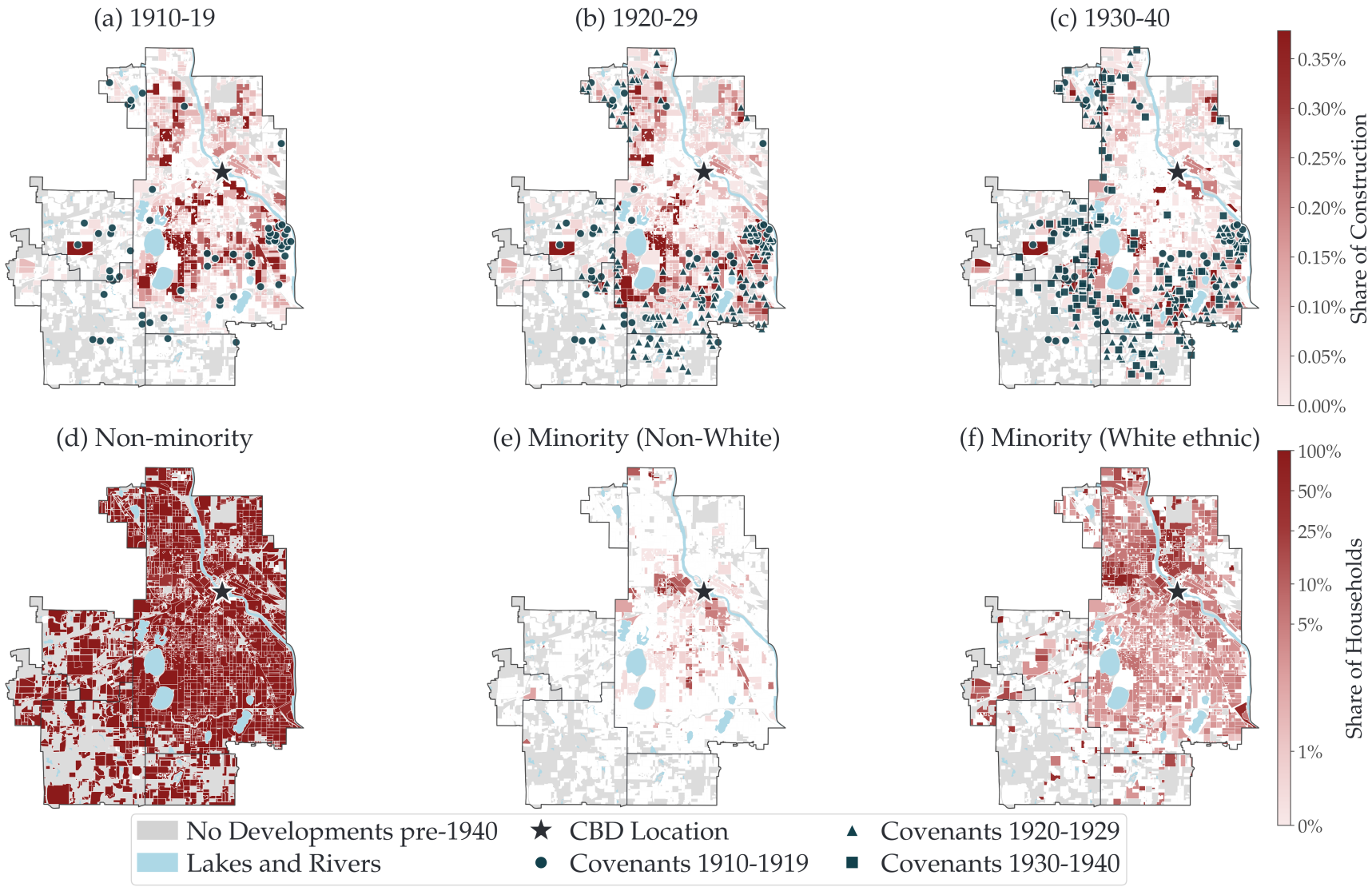
For our analysis, we need to link households, racial covenants, and other data on development characteristics to geolocated developments in 1940. In the 1940 census, we have complete addresses for 75.1% of households. For the 24.9% of households with incomplete addresses, we impute missing street names and house numbers based on the

[7] See Appendix Section A.2 and Appendix Figure A.1 for details on this imputation process.

[8] For 5 developments lacking lot size data, we impute the missing values using the mean log lot size from their respective EDs.

[9] For 1,079 developments with missing build years and 19,185 households, we impute missing years with the mean build year of their ED. If that is also missing, we use the municipality's mean build year.

Figure 1: Expansion of Housing Construction and Racial Covenants over Time



Note: Subfigures (a)- (c) plot the percentage of a decade's overall housing construction which took place in each development in red. The blue circles, triangles, and squares indicate where a racial covenant is present in a development from the 1910s, 1920s, and 1930s respectively. Subfigures (d)- (f) plot the share of non-minority, non-White, and White-ethnic households within a development in 1940. See Section 2.4 for the demographic categories. The city boundary of Minneapolis, St. Louis Park, Edina, Robbinsdale, Richfield, and Hopkins is marked in black. Grey areas represent developments with no housing construction by 1940. The black star is the location of the central business district (CBD). Lakes and rivers are marked in blue.

enumeration order, as enumerators collected the census data on the same side of the street in order (Logan and Parman, 2017). Since this order is preserved in the census, we are largely able to impute missing house numbers and street names (see Appendix Table A.2), giving us complete addresses for 96.1% of the 1940 households in the Minneapolis metro area. We match these households with complete addresses to the geolocated historical developments. For the remaining 3.9% of households with missing addresses, we randomly assign these unmatched households across developments within their respective EDs.[10] Additionally, we link racial covenants, developers, and other geographic data to the historical developments. Further details on the geo-linking across datasets used to construct the final household-development dataset can be found in Appendix A.

Figure 1 a)–c) plot the distribution of new housing construction by development in three different periods–1910-19, 1920-29, 1930-40. The percentage of new construction within a development by decade is color-coded in green; the darker the color, the higher the construction. As the figure shows, the construction is expanding radially outwards over time from the central business district (CBD), indicated by the black star. Additionally, the figure shows the introduction of racial covenants by decade. As demonstrated, the use of racial covenants began to spread continuously after 1910, the year when the first racial covenant was established, with a notable increase after 1930.

Neighborhoods: Our definition of neighborhoods is based on EDs and historic real estate development. We focus on the development level because it is at this level that developers constructed groups of similar houses and made decisions about whether to include racial covenants and which households would be admitted into the development. Development size varies: the 25th-percentile development contains 4 households, the 75th-percentile development contains 48, and the mean contains 52.2 (see Appendix Table A.3). We begin with developments as the fundamental building block of a neighborhood but adjust our definitions at ED boundaries when necessary. Approximately 500 developments cross ED boundaries, typically spanning two EDs and rarely more. This results in development–ED intersections that maintain developments as the relevant unit

[10] In Appendix B, we show that our findings are not determined by the random placement of these 3.9% of households.

for household location choices and admission rules while allowing characteristics to vary across EDs. For broader spatial fixed effects, we combine 4 to 12 contiguous EDs into "cluster EDs" that have broadly comparable populations; a cluster ED contains roughly a median of 9,000 residents.[11] Therefore, our final definition of a neighborhood is the development–ED intersection. Further details can be found in Appendix Section A.11.

2.4 Historical Minority Groups and Evidence from Racial Covenants

In 1940, the demographic groups that could have faced neighborhood-choice restrictions were not defined solely by race, as it is often measured today. Following the work of real estate economist Homer Hoyt, who ranked neighbors based on racial and ethnic groups in 1933 (Hoyt, 1933), we categorize demographics into three groups: non-White households (regardless of birthplace), White-ethnic households (comprising White first-generation immigrants from Southern and Eastern Europe, Asia, Africa, and Central and South America), and White non-minority households (consisting of U.S.-born White households and White immigrant households from Western and Northern Europe) (see Appendix Figure B.1).[12] Historical accounts indicate Scandinavians' extended influence in the economy and politics of the Minneapolis metro area (Anderson and Blanck, eds, 2001; Bergman, 2017). In our 1940 sample, the first two groups—non-White and White-ethnic households—account for 7.6% of households.

We corroborate this categorization by examining whether racial covenants targeted the racial and ethnic minorities identified by Hoyt. We employ an instrumental variable strategy because developers did not choose where to add covenants randomly. We create an instrument for covenanted neighborhoods called the urbanization frontier index, which measures the share of nearby buildable lots that had already been developed toward the city center by 1930. This index reflects the outward expansion of development in the metro area during a time when the adoption of covenants was increasing.

The relevance condition relies on the correlation between this outward expansion into

[11] Since EDs were created primarily for administrative purposes rather than for economically or socially meaningful neighborhoods and varied significantly in population (Potter, 2012), the number of EDs combined for cluster EDs varies accordingly.

[12] Similar ethnic hierarchies appeared in contemporary eugenic writing. Brigham's (Brigham, 1923) purported intelligence rankings broadly favored Northern and Western European origins over Southern and Eastern European origins (see Appendix Figure B.2).

virgin and agricultural land and the rising trend of developers adding racial covenants to new real estate projects that peaked during the 1930s. Consequently, developments closer to the urbanized frontier of the metro area were more likely to include racial covenants in 1940 than those closer to the city center. The exclusion restriction assumes that, conditional on building along the frontier, the exact location of development was effectively random. In this context, as farmland was converted into urban areas, land available for development often arose from farm turnover and projects influenced by unpredictable financial changes experienced by developers. Further details about the instrument and empirical strategy can be found in Appendix Section B.1.

Using this instrumental variable approach, we find that residing in a fully covenanted neighborhood decreases the likelihood of a non-White household by 10 percentage points (Appendix Table B.1). Additionally, we show that these covenants also affected White ethnic minorities. Specifically, living in a fully covenanted neighborhood reduced the probability of a White Eastern or Southern European immigrant household residing there by 2.8 percentage points. Figure 1d–f shows the spatial distribution of the three groups; Appendix Figure B.5 shows the distribution of minority shares across developments.

These findings show how *explicit* discrimination impacted household sorting in 1940, thereby limiting neighborhood choices for both non-White and White ethnic households, which we collectively refer to as minority households. The sorting results are consistent with the idea that unequal access affected non-White and some White-ethnic households; however, they cannot, on their own, distinguish restrictions from household preferences or the effects of prices or other location characteristics. The model discussed in Section 3 distinguishes these channels and also allows for unobserved *implicit* neighborhood choice set restrictions. This evidence motivates demographically differentiated admission rules in the model.

3. Model

We build a two-sided matching model of selective admissions, following Gandhi (2026) and Agarwal and Somaini (2025). The model has two parts. On the demand side, households choose a neighborhood according to a random utility model. On the supply side,

a reduced-form admission rule determines which neighborhoods admit each household, based on the household's demographic group. The model departs from standard location choice models in two ways. First, choice sets differ across demographic groups. Second, choice sets are unobserved, or latent, and we estimate them from the data. Observed sorting therefore reflects both who wants to live in a neighborhood and who is permitted to live there.

3.1 A Two-Sided Matching Model of the Housing Market

3.1.1 Demand

Household i of demographic group $k(i)$ chooses among the neighborhoods in its choice set $\mathcal{O}_i \subseteq \mathcal{J}$. Households belong to one of three groups, $k(i) \in \{nm, we, nw\}$: non-minority, White-ethnic, and non-White households (Section 2.4). The indirect utility of household i in neighborhood j is

$$U_{ij} = \delta_j + \Delta\beta^{k(i)} X_j^d + \epsilon_{ij}, \qquad \epsilon_{ij} \sim \text{Type I extreme value.} \tag{1}$$

The term $\Delta\beta^{k(i)} X_j^d$ is the utility difference of group $k(i)$ relative to non-minority households, the reference group, so that $\Delta\beta^{nm} = 0$.[13] The vector of demand covariates is

$$X_j^d = \left[\log \text{price}_j,\ \text{non-minority share}_{j\text{-ring}},\ \log \text{lot size per household}_j\right].$$

We measure the non-minority share over the nearest neighborhoods, by centroid distance, that together hold at least 500 households, including neighborhood j itself. The term δ_j is the component of utility common to all groups, and ϵ_{ij} is an idiosyncratic shock, i.i.d. across households and neighborhoods. The non-minority valuation of the covariates enters through δ_j:

$$\delta_j = \beta^{nm} X_j^d + W_j'\gamma + \xi_j,$$

where W_j collects characteristics that all groups value equally: distances to the central business district, streetcar lines, heavy industry, commercial areas, and the nearest lake,

[13] This reference-group parameterization is without loss. The total coefficient of group k is $\beta^k \equiv \beta^{nm} + \Delta\beta^k$, which Table 1 reports.

log development size, and area fixed effects. The term ξ_j is a neighborhood-level demand shock that households observe but the econometrician does not.

Each household can also choose an outside option, $j = 0$, which represents living elsewhere in Minnesota and has utility $U_{i0} = \epsilon_{i0}$. We assume that the outside option belongs to every choice set. Conditional on her choice set, the probability that household i chooses neighborhood $j \in \mathcal{O}_i$ is

$$\mathbb{P}_{ij}(\Delta\beta, \delta \mid \mathcal{O}_i) = \frac{\exp\left(\delta_j + \Delta\beta^{k(i)} X_j^d\right)}{1 + \sum_{j' \in \mathcal{O}_i} \exp\left(\delta_{j'} + \Delta\beta^{k(i)} X_{j'}^d\right)}, \qquad \mathbb{P}_{ij} = 0 \text{ if } j \notin \mathcal{O}_i. \tag{2}$$

3.1.2 Admission Rule

Neighborhood j admits household i of group $k(i)$ if and only if

$$\alpha^{k(i)} X_j^s \geq \eta_i, \qquad \eta_i \sim \mathcal{N}(0,1), \tag{3}$$

where $\alpha^{k(i)}$ are group-specific admission parameters and η_i is an idiosyncratic household draw, common across all neighborhoods, that is unknown to the econometrician, such as i's accumulated wealth. The admission probability is therefore

$$\mathbb{P}\left(j \in \mathcal{O}_i\right) = \Phi\left(\alpha^{k(i)} X_j^s\right).$$

The vector of admission covariates is

$$X_j^s = \left[\text{developer's leniency}_j,\ \text{covenant share}_j,\ \text{log price}_j,\ \text{log development size}_j\right],$$

together with a group-k constant, capturing a common component of group-k that affects the admissions to all neighborhoods.[14] Developer's leniency is an indicator for whether the neighborhood's developer has an above-median minority share in its other devel-

[14] Covenants are primarily added by the developer at the time of registering or subdividing the development and may be a response to the preferences of the households that form demand. In this paper, we take the covenant share of each development as an exogenously determined outside our model.

opments.[15] These covariates proxy for four potential channels of restriction: developer-specific screening through leniency, explicit discrimination through covenants, financial constraints through prices, and differences in admission related to scale through development size.

Because the draw η_i is common across neighborhoods, choice sets are nested. Order the neighborhoods by their admission index for group k, so that $\alpha^k X^s_{(1)} \geq \cdots \geq \alpha^k X^s_{(J)}$. A household admitted to neighborhood (n) is also admitted to every neighborhood with a higher index. Therefore, each household's choice set is therefore one of the $J+1$ sets $O^n \equiv \{(1), \ldots, (n)\}$, $n = 0, \ldots, J$, with probability

$$\mathbb{P}\big(O_i = O^n\big) = \Phi\big(\alpha^{k(i)} X^s_{(n)}\big) - \Phi\big(\alpha^{k(i)} X^s_{(n+1)}\big),$$

where $\Phi(\alpha^k X^s_{(0)}) \equiv 1$ and $\Phi(\alpha^k X^s_{(J+1)}) \equiv 0$.[16] A household with the empty choice set, $n = 0$, is admitted nowhere and chooses the outside option. The model with a common draw η_i is convenient for computational reasons. It reduces the match probability to a sum over $J+1$ nested choice sets with closed-form weights. This avoids both numerical integration and a sum over all 2^J subsets of $\mathcal{J}$. Appendix C.1 extends the model to admission draws that vary across neighborhoods, η_{ij}.

3.2 Ignoring Access Restrictions

If the econometrician ignores admission and assumes $O_i = \mathcal{J}$ for every household, the model collapses to the unrestricted logit

$$\widetilde{\mathbb{P}}_{ij}(\Delta\beta, \delta) = \frac{\exp\big(\delta_j + \Delta\beta^{k(i)} X^d_j\big)}{1 + \sum_{j' \in \mathcal{J}} \exp\big(\delta_{j'} + \Delta\beta^{k(i)} X^d_{j'}\big)}. \tag{4}$$

This specification is misspecified whenever choice sets do not expand the full set of neighborhoods. It attributes low demand for hard-to-access neighborhoods to preferences

[15] Developer leniency is missing for 59% of locations and 40% of households for two reasons. First, many developers built a single development, so they have no other developments from which to measure leniency. These account for 25% of developments and 25% of households. Second, we cannot identify the developer of 36% of developments. We include a separate indicator for those locations with missing leniency in the admission rule.

[16] See Appendix C.1 for details.

rather than to admission restrictions, and therefore confounds tastes with access. This is the same latent selection logic emphasized by Dubé et al. (2021): a latent indicator switches an alternative into or out of the effective choice set.

3.2.1 A Toy Model of Restricted Access

To isolate the source of the bias, we consider a simplified environment. Household i gets utility from neighborhood $j \in \mathcal{J}$

$$U_{ij} = \beta_0 X_j + \epsilon_{ij}, \qquad \epsilon_{ij} \overset{\text{iid}}{\sim} \text{Type I extreme value},$$

where X_j is a neighborhood characteristic, such as price or lot size, and β_0 is the true taste for X_j. Let $O_i \subseteq \mathcal{J}$ be the set of neighborhoods that admit household i. Suppose that the econometrician ignores the restrictions and fits a logit on the full set $\mathcal{J}$, and let $\widetilde{\beta}$ be the probability limit of her estimate. For any choice set $\mathcal{S}$, define the logit-weighted average of the characteristic X as

$$m_{\mathcal{S}}(\beta) \equiv \frac{\sum_{j \in \mathcal{S}} X_j \, e^{\beta X_j}}{\sum_{j \in \mathcal{S}} e^{\beta X_j}}.$$

Define the *selection gap* as

$$\Gamma \equiv \underbrace{\mathbb{E}_i\left[m_{O_i}(\beta_0)\right]}_{\text{average chosen } X \text{ under actual access}} - \underbrace{m_{\mathcal{J}}(\beta_0)}_{\text{average chosen } X \text{ with full access}}.$$

Theorem 1 (Ignoring restrictions biases preference estimates)**.** *The pseudo-true coefficient $\widetilde{\beta}$ uniquely solves $m_{\mathcal{J}}(\widetilde{\beta}) = \mathbb{E}_i\left[m_{O_i}(\beta_0)\right]$, and*

$$\widetilde{\beta} - \beta_0 = \frac{\Gamma}{\text{Var}_w(X_j)} \qquad \Longrightarrow \qquad \text{sign}\left(\widetilde{\beta} - \beta_0\right) = \text{sign}(\Gamma),$$

where $\text{Var}_w(X_j)$ is the variance of X_j under logit weights on $\mathcal{J}$, evaluated at a coefficient between $\widetilde{\beta}$ and β_0.

The proof is in Appendix C.2. The misspecified logit matches one moment: the average X of chosen neighborhoods. It cannot separate access from taste, so it attributes the selection gap Γ to taste.

Corollary 1. *Restrictions that remove high-X_j (low-X_j) neighborhoods from households' choice sets bias the estimated taste downward (upward).*

This result maps naturally to financial constraints in housing markets. For example, if X_j is the price of a home, or a neighborhood characteristic strongly correlated with price, then a down-payment constraint can make all homes above a household's liquidity threshold infeasible.[17] A household with liquid wealth b_i can only purchase homes whose required down payment d_j satisfies $d_j \leq b_i$. If d_j is increasing in X_j, the household's choice set has the form $O_i = \{j : X_j \leq c_i\}$. A model that treats high down-payment homes as available then mistakes the absence of choices from those homes for weaker demand, rather than for lack of access. In Monte Carlo simulations, the bias can be large enough to flip the sign of the estimated taste (Appendix C.3).

3.2.2 From the Toy Model to the Full Model

In the full model, an unrestricted demand model still tries to rationalize the observed distribution of chosen neighborhoods using only preferences. When the econometrician assumes $O_i = \mathcal{J}$ for every household, choices generated by both preferences and admission are explained only through δ_j and the demand coefficients. If admission probabilities are systematically related to prices, neighborhood demographic composition, covenants, or other neighborhood characteristics, a model without admission absorbs this relation into the estimated demand parameters.

The toy model gives a clean sign prediction because there is only one characteristic and one coefficient. With several demand characteristics, the sign of any individual coefficient can also depend on how the characteristics covary with one another. Appendix C.2 states the corresponding multivariate result. Two further features of the full model can change the size and the sign of the bias. First, the mean utilities δ_j are shared across groups and estimated jointly with the demand coefficients. The restriction then does not disappear into δ_j. It distorts the estimated δ_j as well. Second, the outside option belongs to every choice set. A household admitted nowhere chooses the outside option, so restrictions also reduce the share of each group that lives inside the metro area. A model without admission must rationalize this lower inside share through preferences, which can move the estimated

[17] Note that typical down-payments in Minneapolis were in the range of 40% to 50% in 1940.

coefficients in the direction opposite to the toy model. Section 4.4 compares the estimates of the full model with those of models that ignore or simplify access restrictions.

4. Estimation and Results of the Two-Sided Matching Model

This section describes how we estimate the model of Section 3. It then discusses identification and presents the estimates.

4.1 Match Probabilities

We observe where each household lives but not its choice set. The probability of an observed match therefore integrates over admission and choice:

$$\mathbb{P}\big(y_{ij}=1 \mid k(i);\alpha,\Delta\beta,\delta\big)=\sum_{O}\underbrace{\mathbb{P}\big(O_i=O;\alpha\big)}_{\text{admission}}\times\underbrace{\mathbb{P}_{ij}\big(\Delta\beta,\delta \mid O\big)}_{\text{demand}}, \tag{5}$$

where $y_{ij}=1$ if household i lives in neighborhood j. Let $a_j^k \equiv \alpha^k X_j^s$ denote the admission index of neighborhood j for group k. Let $v_j(k) \equiv \exp\left(\delta_j + \Delta\beta^k X_j^d\right)$ denote its demand index. With nested choice sets, only the sets O^n that contain j contribute to the sum. Let $\text{rank}_k(j)$ be the position of j when we order neighborhoods by a_j^k from highest to lowest. Then

$$\mathbb{P}\big(y_{ij}=1 \mid k\big)=\sum_{n\geq \text{rank}_k(j)}\Big[\Phi\big(a^k_{(n)}\big)-\Phi\big(a^k_{(n+1)}\big)\Big]\frac{v_j(k)}{1+\sum_{\ell\leq n} v_{(\ell)}(k)}. \tag{6}$$

The probability of the outside option is $\mathbb{P}(y_{i0}=1 \mid k)=1-\sum_{j\in\mathcal{J}}\mathbb{P}(y_{ij}=1 \mid k)$. It includes households admitted nowhere.

4.2 Estimation Procedure

We estimate the model in two steps, extending Goolsbee and Petrin (2004) and Bayer et al. (2007) to latent choice sets. The first step recovers $(\alpha,\Delta\beta)$ by maximum likelihood. It also recovers the mean utilities δ_j through a contraction mapping. The second step projects δ_j on neighborhood characteristics with instrumental variables. This step recovers the preferences of non-minority households, β^{nm}, and the coefficients γ.

First step. For given $(\alpha,\Delta\beta)$, we evaluate the likelihood as follows:

1. For each group k with an admission rule, compute a_j^k and order neighborhoods by it. This ordering gives the nested choice sets O^n and their probabilities $\Phi(a^k_{(n)})$ –

$\Phi(a^k_{(n+1)})$.

2. Recover δ with a modified version of the contraction mapping in Berry et al. (1995).[18] Let s_j be the observed share of households in the market who live in j. Let N^k be the number of group-k households in the market and $N = \sum_k N^k$. Start from a guess δ^0. At iteration t:

   (a) Compute $\mathbb{P}(y_{ij} = 1 \mid k; \delta^t)$ from (6) for every group.

   (b) Compute the predicted share $S_j(\delta^t) = \sum_k \frac{N^k}{N} \mathbb{P}(y_{ij} = 1 \mid k; \delta^t)$.

   (c) Update $\delta_j^{t+1} = \delta_j^t + \log s_j - \log S_j(\delta^t)$.

   Repeat until convergence. Denote the result by $\delta(\alpha, \Delta\beta)$.

3. Compute the log-likelihood

$$\mathcal{L}(\alpha, \Delta\beta) = \sum_k \sum_{j \in \mathcal{J} \cup \{0\}} N_j^k \log \mathbb{P}\big(y_{ij} = 1 \mid k; \alpha, \Delta\beta, \delta(\alpha, \Delta\beta)\big),$$

   where N_j^k is the number of group-k households in neighborhood j. N_0^k is the number of group-k households elsewhere in Minnesota.

We maximize $\mathcal{L}$ with a Nelder–Mead search followed by a Newton trust-region step, from several starting values. The contraction matches the total share of each neighborhood. The variation in group composition across neighborhoods, together with each group's share in the metro area, identifies α and $\Delta\beta$.

Second step. The first step delivers $\hat{\delta}_j = \delta_j(\hat{\alpha}, \widehat{\Delta\beta})$. We then estimate

$$\hat{\delta}_j = \beta^{nm} X_j^d + W_j' \gamma + \xi_j \tag{7}$$

by two-stage least squares. Observations are locations, weighted by their number of households. W_j includes coarse ED fixed effects. Section 4.3.3 describes the instruments.

[18] Berry et al. (1995) show that this mapping is a contraction in the standard logit. We do not have a proof for the model with admission. We therefore check convergence at every evaluation of the likelihood.

4.3 Identification

The first step identifies the admission parameters α, the group differences in preferences $\Delta\beta$, and the mean utilities δ_j. The second step identifies the non-minority preferences β^{nm} and γ. We first describe how we choose which groups face selective admissions. We then discuss each step.

4.3.1 Testing for Selective Admissions

Before estimating the model, we determine which demographic groups face selective admissions. Under our admission rule, the draw η_i has unbounded support. Hence, for any finite admission index,

$$\mathbb{P}\big(j \in O_i \mid k(i) = k\big) = \Phi\big(\alpha^k X_j^s\big) < 1.$$

Full admission is therefore a limit and requires the admission index to diverge for every neighborhood. Consequently, if a group faces no admission barriers, an estimated admission rule pushes its parameters toward this limit and the optimizer may not converge and behave poorly.

Therefore, before running the full model, we test for selected admissions as follows. For each group k, we compare full admission with a model that estimates its admission rule while holding the treatment of the other groups fixed:

$$H_0 : \ \mathbb{P}\big(j \in O_i \mid k(i) = k\big) = 1 \ \text{ for all } j, \qquad H_1 : \ \alpha^k \text{ estimated}, \qquad LR = 2(\ell_1 - \ell_0),$$

where ℓ_0 and ℓ_1 are the maximized log-likelihoods under the two specifications. Because the null lies on the boundary of the parameter space, namely $\alpha^k \to \pm\infty$, the LR statistic therefore does not follow its usual χ^2 distribution. We compute p-values by parametric bootstrap under the fitted null.

We begin with full admission for all three groups. At each step, we consider adding an admission rule for each remaining group. We test the group with the largest likelihood gain. If we reject full admission for that group at the 5% level, we keep its admission rule and repeat. Otherwise, we stop.

The procedure first selects selective admissions for the two minority groups, non-White and White-ethnic households. For non-minority households, the test delivers that they are admissible everywhere. When every group has an estimated admission rule, no group anchors the level of admission, and this level is not identified. We therefore impose full admission for non-minority households. Two facts support this choice. First, non-minority households live in 98.6% of neighborhoods, against 49.5% for White-ethnic and 8.5% for non-White households. Second, the likelihood gain from an admission rule is at least five times smaller for non-minority households than for either minority group.

4.3.2 First Step: Admission Rules versus Preferences

We first need several normalizations. The level of utility is not identified, so we normalize the utility of the outside option to $U_{i0} = \epsilon_{i0}$ (Train, 2009). The scale of utility is not identified either, so ϵ_{ij} follows a standard Type I extreme value distribution. On the admission side, we normalize η_i to have mean zero and unit variance. The group constants in X_j^s absorb the level of each group's admission index. Full admission for non-minority households anchors the overall level of admission.

A household lives in a neighborhood only if it wants to live there and the neighborhood admits it. Observed choices mix these two channels. Under our functional-form assumptions, logit errors and a normal admission draw, the model is identified without exclusion restrictions (Appendix D.2). This argument is fragile, as it attributes any departure from logit substitution patterns to the formation of choice sets. We therefore also rely on an excluded shifter (Goeree, 2008; Agarwal and Somaini, 2025). This variable shifts the formation of choice sets but not household utility, conditional on the other covariates.

Our excluded shifter is developer leniency, an indicator for whether the neighborhood's developer has an above-median minority share in its *other* developments.[19] We assume that developer leniency does not *differentially* enter utility across demographic groups, conditional on the demand covariates: price, the non-minority share, and lot size per household. The mean utility δ_j is specific to each neighborhood. Any valuation of leniency that all groups share is therefore part of δ_j and does not violate the exclusion. The rationale is that developers may have preferences over who lives in their developments,

[19] Appendix D.1 describes the construction of this variable and reports balance tests.

beyond monetary incentives and can shape a development's demographic composition through whom they admit. This shifter follows the logic of judge-leniency designs (Kling, 2006; Dahl et al., 2014; Dobbie et al., 2018). In those designs, a decision maker's choices in other cases predict the decision in the current case. Here, a developer's admission of minority households in its other developments predicts its admission decisions in this development. As in those designs, we leave out the development itself, so the measure does not mechanically reflect its own residents. Unlike in those designs, developers are not randomly assigned to locations. The balance tests in Appendix D.1 therefore stand in for the random-assignment checks of that literature. [20]

The covenant share also enters the admission rule but not the group-specific terms of utility. We do not base our argument on this exclusion. Minority households may have avoided covenanted neighborhoods for reasons other than admission, and developers did not place covenants at random (Section 2.4).

Appendix D.2 shows how the excluded shifter separates the two channels. The argument compares minority with non-minority households in the same neighborhood, which removes δ_j. A change in leniency moves the minority match probability only through admission. A change in price moves it through both admission and utility. The response to leniency reveals how strongly admission responds. Subtracting this admission response from the price response leaves the preference for price.

4.3.3 Second Step

The second step projects $\hat{\delta}_j$ on X_j^d and W_j in (7). The demand shock ξ_j enters δ_j additively. Two covariates in X_j^d are equilibrium outcomes and can depend on ξ_j: the log price and the non-minority share. Ordinary least squares would then give biased estimates. We instrument both.

First, we instrument the log price with terrain slope, following Saiz (2010). Steeper slopes raise construction costs and reduce housing supply. The identifying assumption is that slope affects prices through construction costs but does not affect housing demand directly. This assumption is more plausible in the relatively flat lake geography of the

[20] Lot size per household and the non-minority share enter group-specific utility but not the admission rule.

Minneapolis metro area.

Second, we instrument the non-minority share with a shift-share instrument following Card (2001) and Tabellini (2020). For each ED e, we construct

$$Z_{e,1940} = \frac{M_{e,1930} + \sum_r w_{er}(1930)\, O^{Q}_{r,1930-38}}{\hat{P}_{e,1940}}, \tag{8}$$

where $w_{er}(1930)$ is the share of households from foreign country r who lived in e in 1930. $M_{e,1930}$ is the 1930 stock of minority households in e. $O^{Q}_{r,1930-38}$ is the yearly immigration quota for country r under the US Immigration Act of 1924, multiplied by 9 to cover 1930–1938. The quotas provide national variation in the inflow of households from each country.[21] $\hat{P}_{e,1940}$ is the predicted 1940 population of e. Appendix D.3.1 describes the construction in detail.

The shift-share instrument relies on two assumptions. First, the 1930 settlement shares are not correlated with unobserved demand shocks in 1940. All EDs belong to the same metro area, so similar local economic trends are likely to affect them. Second, the national quotas are exogenous to local shocks, as argued by Tabellini (2020).

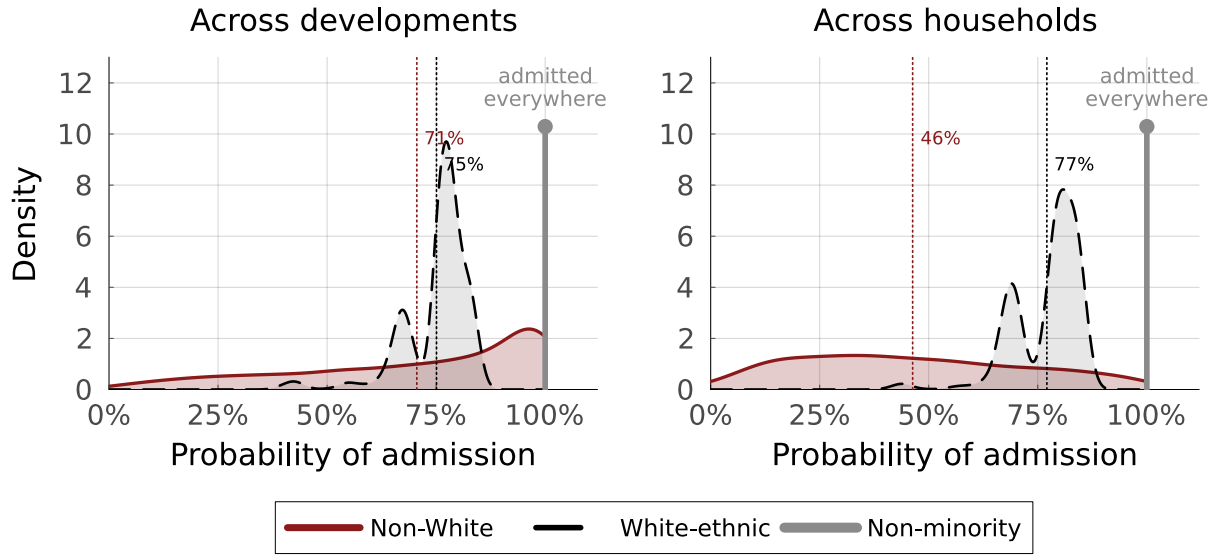
Appendix Table D.2 reports the first-stage regressions. Terrain slope is positively correlated with log rents, consistent with steeper slopes raising construction costs. The shift-share instrument is negatively correlated with the non-minority share. The instruments are jointly weak by conventional standards. The Kleibergen–Paap F statistic is 5.94. The Lewis–Mertens statistic is 19.40, below its critical value of 45.38. We therefore treat the reported standard errors of the second-step coefficients as a lower bound on their uncertainty.

4.4 Results

Admission probabilities differ substantially across demographic groups (Figure 2). Non-minority households are admitted everywhere by assumption (Section 4.3.1). On average across developments, White-ethnic households can access 75% of neighborhoods and non-White households 71%. Non-White access varies widely, from near-full in some

[21] At least 87% of households in the minority groups are White-ethnic, and hence foreign-born. We therefore use a shifter that affects the foreign-born population rather than the US-born population.

Figure 2: Estimated Admission Probabilities by Demographic Group



Note: This figure plots the distribution of estimated admission probabilities for White-ethnic and non-White households. The left panel weights each development equally. The right panel weights developments by their number of households. Vertical lines mark the group means. Non-minority households have full admission in the model.

developments to heavily restricted in others. Non-White households have greater access to smaller developments. Weighted by households, mean non-White access falls from 71% to 46%. The corresponding White-ethnic figure is 77%.

Table 1 presents the admission parameters α^k from Equation (3) and the demand coefficients $\beta^k \equiv \beta^{nm} + \Delta\beta^k$ from Equation (1). Standard errors are in parentheses. Neither minority group had access to every neighborhood in the metro area, but the pattern of restricted access differs between the two groups. A higher share of racial covenants lowers the admission probability for both groups. Developer leniency raises admission for both groups. Price and development size matter only for non-White households. Higher-priced and larger developments are less likely to admit non-White households. For White-ethnic households, neither coefficient is statistically different from zero. Restrictions therefore extend beyond those recorded in racial covenants. A focus on explicit restrictions alone would miss important differences in access.

Turning to demand, the price coefficients are negative for all three groups. The minority coefficients are larger in magnitude than the non-minority coefficient, so minority

Table 1: Estimation Results

Panel A. Demand

	Two-sided matching model (1)	Only covenants for exclusion (2)	Standard model without restrictions (3)
Non-minority β^{nm}			
Rent	−0.236 (0.165)	−0.323* (0.167)	−0.323* (0.167)
Non-min. share	0.380** (0.148)	0.392*** (0.141)	0.392*** (0.141)
Lot size per HH	−0.031 (0.038)	−0.008 (0.038)	−0.008 (0.038)
White-ethnic β^{we}			
Rent	−0.369** (0.168)	−0.398** (0.168)	−0.425** (0.168)
Non-min. share	−0.268* (0.148)	**−0.162** (0.141)	**−0.168** (0.141)
Lot size per HH	0.061 (0.042)	**0.144***** (0.042)	**0.148***** (0.041)
Non-White β^{nw}			
Rent	−1.125*** (0.176)	−1.099*** (0.171)	−1.131*** (0.171)
Non-min. share	−0.443*** (0.149)	**−0.198** (0.142)	**−0.197** (0.142)
Lot size per HH	−0.286*** (0.049)	**−0.142***** (0.050)	**−0.126**** (0.049)
Locations in second step	3,937	3,883	3,883
p: all coefficients equal to two-sided	—	< 0.001	< 0.001

Panel B. Admission rule

	Non-minority	White-ethnic	Non-White
Rent	—	−0.009 (0.064)	−0.637*** (0.108)
Covenant share	—	−0.131*** (0.028)	−0.331** (0.131)
Developer leniency	—	0.193*** (0.027)	0.288*** (0.036)
Development size	—	0.057 (0.036)	−1.088*** (0.082)

Panel C. Elasticities

	Non-minority	White-ethnic	Non-White
Elasticity, demand only	−0.244	−0.382	−1.159
Elasticity, demand + admission	−0.244	−0.385	−1.527
Locations (households)		3,937 (164,669)	

Notes: Standard errors in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Bold estimates in columns (2)–(3) differ from the corresponding two-sided estimate at the 5% level. Demand coefficients are group-specific totals; their standard errors combine the level and gap variances, assuming zero covariance between steps. Standard errors for imposed choice sets are conditional on the estimated location effects. Non-minority share is measured over the nearest locations, by centroid distance, that together hold at least 500 households, including the location itself. Covariates are standardized. The two-sided model uses 164,669 households in 3,937 locations. Positive admission coefficients increase admission probabilities. Admission intercepts and the indicator for a missing developer propensity are estimated but not reported.

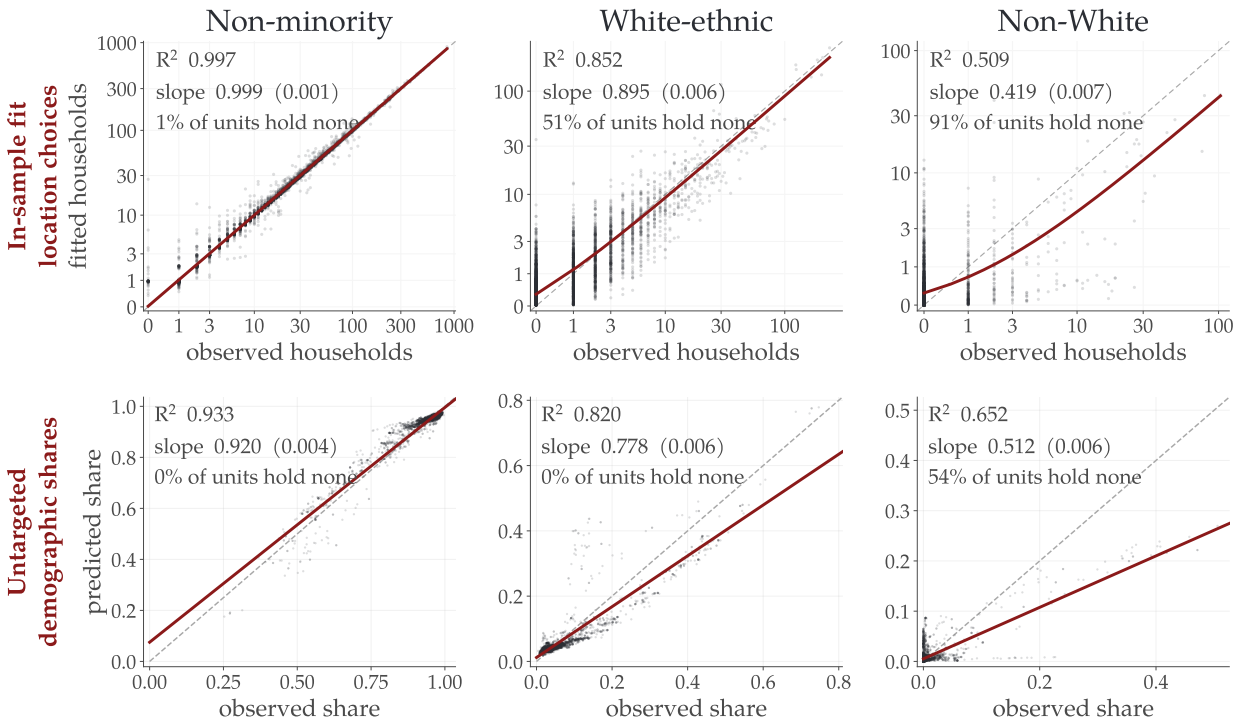
households are more price-sensitive. This finding aligns with previous research (Bayer et al., 2007; Almagro et al., 2026). The non-minority coefficient is not statistically different from zero. The estimates also show homophily. Non-minority households prefer developments surrounded by a higher share of non-minority residents. Both non-White and White-ethnic households prefer developments with a lower share. The White-ethnic coefficient is significant at the 10% level. Non-White households also have a negative coefficient on lot size per household. This preference for smaller houses does not stem from differences in housing costs or from the access restrictions in the model.

Price affects location choices through both demand and admission. For White-ethnic households, the admission coefficient on price is close to zero. Their combined price elasticity is therefore almost equal to the demand-only elasticity (-0.385 against -0.382). For non-White households, higher prices lower both utility and admission. Their combined elasticity is more negative than the demand-only elasticity (-1.527 against -1.159). The elasticity for non-minority households is -0.244, because this group has full admission.

Table 1 also reports two alternative models. The first treats covenants as the only source of restrictions. The second assumes unrestricted access. Neither recovers the demand estimates of the two-sided matching model. In both, the non-minority price coefficient is more negative, although the difference is not statistically significant. The alternative models also imply that White-ethnic households prefer larger lots, with positive and significant coefficients. Once we account for choice-set restrictions, this coefficient cannot be distinguished from zero. The alternative models also weaken White-ethnic households' preference for living near other minority households.

For non-White households, the price coefficients in the alternative models are close to the two-sided estimate. These models, however, understate non-White households' preference for living near other minority households. They also weaken the negative coefficient on lot size. Considering only covenants, or ignoring restrictions altogether, leaves variation in access unmodeled. This unmodeled variation distorts the estimated preferences.

Figure 3: Model Fit



Note: The top row compares observed and fitted household counts by location for each group, on a log scale. The bottom row compares observed and predicted group shares over each location's 500-household ring, at observed prices and composition. The estimation does not target these shares. Maroon lines are fitted lines, with standard errors in parentheses. Statistics are household-weighted.

4.5 Model Fit

Figure 3 compares observed and model-predicted household counts across neighborhoods for each demographic group. The estimation does not require an exact match of each group's count in every neighborhood. The model fits non-minority and White-ethnic counts closely, with R^2 values of 0.997 and 0.852. Fit is weaker for non-White households ($R^2 = 0.509$), particularly where observed counts are low. This is the smallest group, and 91% of neighborhoods have no non-White households. This leaves little variation with which to fit its neighborhood-level counts. The bottom row of Figure 3 reports group shares over each location's ring, which the estimation does not target. The R^2 values are 0.933, 0.820, and 0.652 for non-minority, White-ethnic, and non-White households.

5. Counterfactual Analysis

Table 2: Counterfactual Outcomes

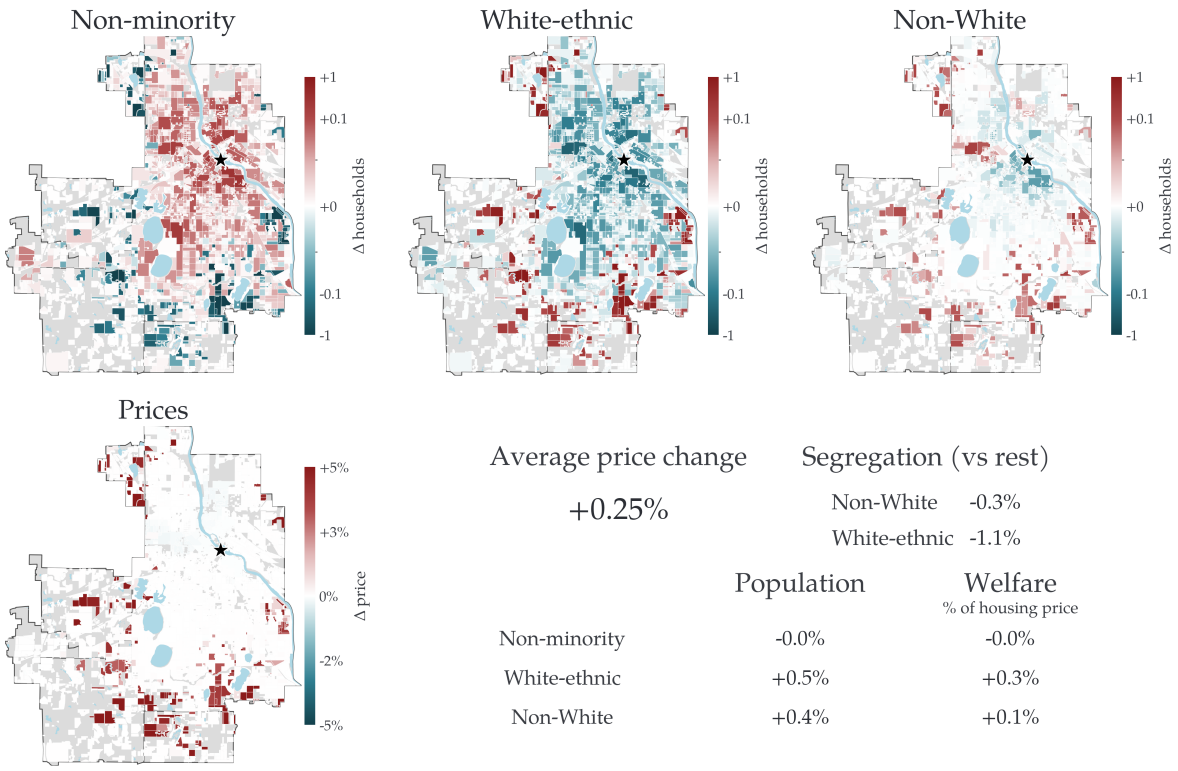
	No recorded covenants	Open access	Open access & no homophily
	(1)	(2)	(3)
Housing prices			
Mean change (geometric)	+0.3%	+9.4%	−1.6%
SD of change in log price	0.011	0.068	0.975
Segregation (adjusted dissimilarity), change			
Non-White	−0.3%	−16.4%	−66.3%
White-ethnic	−1.1%	−10.8%	−96.5%
Households in the metro area, change			
Non-minority	0.0%	−1.8%	−0.2%
White-ethnic	+0.5%	+19.4%	+1.1%
Non-White	+0.4%	+39.3%	+11.6%
All minority	+0.5%	+22.0%	+2.5%
Welfare (compensating variation, % of price)			
Non-minority	0.0%	−1.9%	−0.2%
White-ethnic	+0.3%	+12.4%	−1.0%
Non-White	+0.1%	+13.5%	+2.5%

Note: This table reports changes relative to the model baseline under three counterfactual scenarios. Column (1) sets the covenant share of every development to zero. Column (2) admits every group to every development. Column (3) also removes preferences over neighborhood demographic composition. Rents clear each location against its observed housing stock. Neighborhood composition in demand is held at its observed level. Rent changes are geometric means. Segregation is the adjusted dissimilarity index of Carrington and Troske (1997), computed at the ED level for each group against all other households. Welfare is the compensating variation as a percentage of each group's baseline rent.

This section quantifies the contribution of different mechanisms to the observed segregation patterns. We simulate three equilibrium scenarios. In the first, we remove racial covenants by setting the covenant share of every development to zero. In the second,

we remove all choice-set restrictions, so every group is admitted to every development. In the third, we also remove preferences over neighborhood demographic composition. In all three scenarios, rents adjust to clear the market. We assume a fixed and perfectly inelastic housing supply. Neighborhood demographic composition in the demand model stays at its observed level. We report the geometric mean of rent changes. We measure segregation with the dissimilarity index adjusted for small units (Carrington and Troske, 1997). We compute it at the ED level for each minority group against all other households. In the model baseline, the adjusted index is 0.481 for non-White households and 0.287 for White-ethnic households. The corresponding values in the data are 0.691 and 0.379. Table 2 summarizes the results.

Figure 4: Counterfactual 1: No Recorded Covenants



Note: This figure maps the change in the number of households of each group and the change in rents across developments when the covenant share of every development is set to zero. The bottom-right panel reports the average rent change, the change in segregation for each minority group against all other households, and the change in population and welfare for each group. The star marks the central business district.

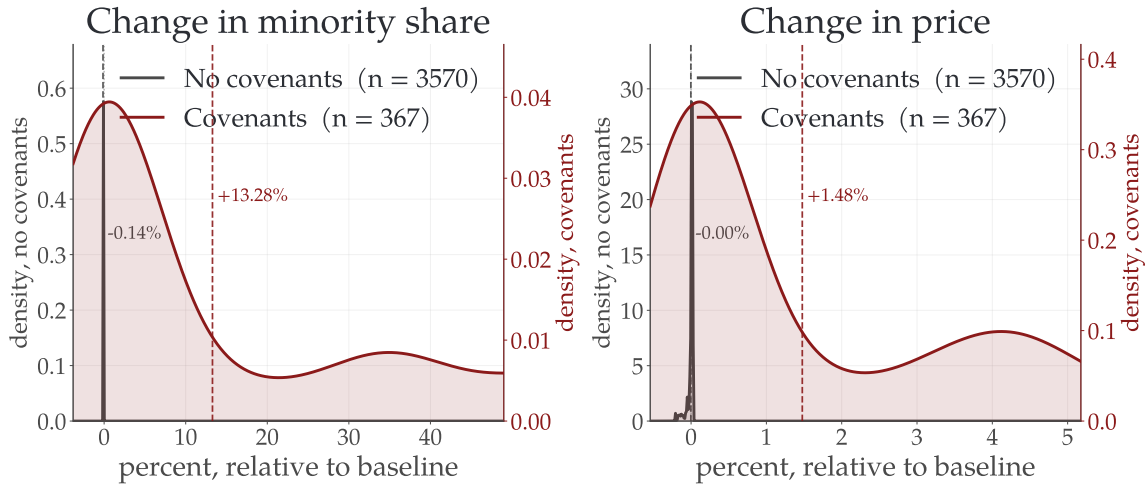
In the first counterfactual, we remove racial covenants. This would be equivalent to

Table 3: Explicit versus Implicit Restrictions

	Baseline	No recorded covenants	No admission restrictions	Covenants' share of the total
Adjusted index of dissimilarity				
Non-White	0.481	0.480	0.402	1.8%
White-ethnic	0.287	0.283	0.256	10.2%
Households in the metro area				
Non-White	1,646	1,652	2,293	0.9%
White-ethnic	10,898	10,949	13,013	2.4%
Housing prices				
Mean change	—	+0.25%	+9.40%	2.7%

Note: This table compares the model baseline with two counterfactuals. "No recorded covenants" sets the covenant share of every development to zero. "No admission restrictions" admits every group to every development. The last column reports the change under "No recorded covenants" as a share of the change under "No admission restrictions". Segregation is the adjusted dissimilarity index at the ED level, for each group against all other households.

Figure 5: Local Effects of Removing Covenants

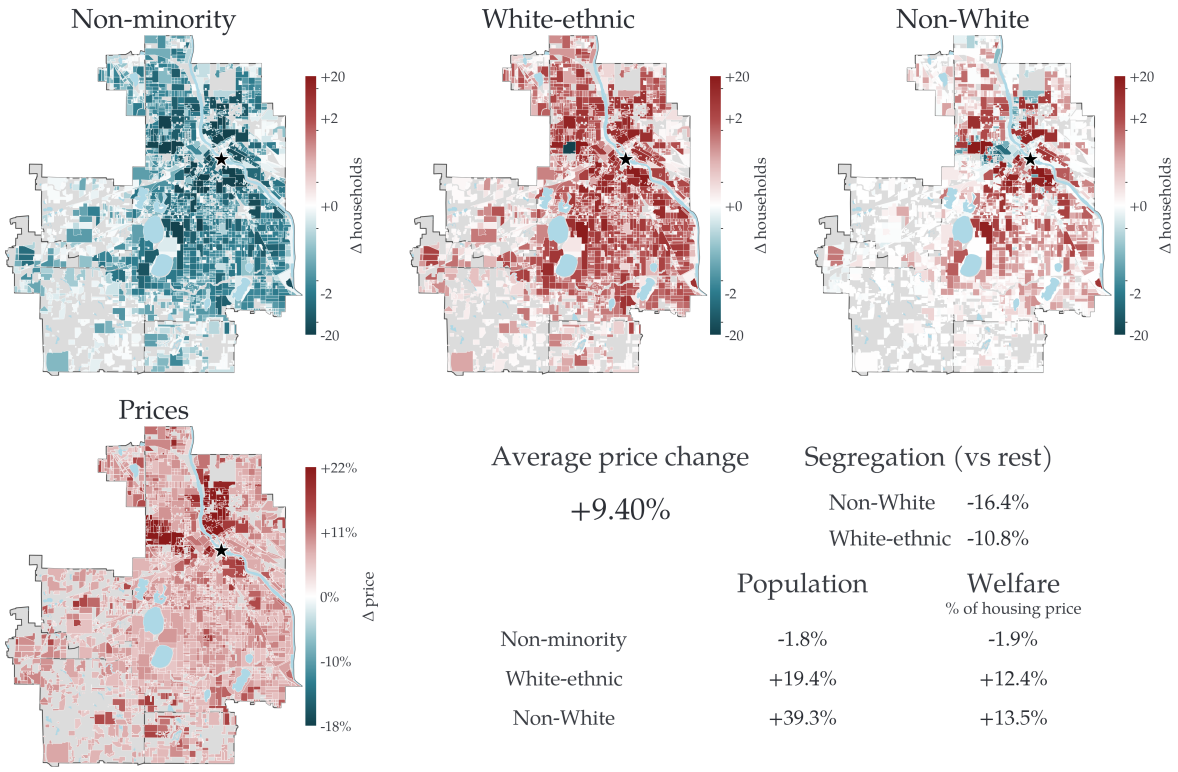


Note: This figure plots the distribution of the change in the minority share (left) and in rents (right) across locations when the covenant share of every development is set to zero. Locations with covenants (367) and without covenants (3,570) are plotted separately, on separate density axes. Dashed lines mark the means.

racial covenants being hypothetically outlawed in 1940. The aggregate effects are small (Figure 4). Segregation falls by 0.3% for non-White households and by 1.1% for White-

ethnic households. The non-White change is close to its simulation standard error of 0.25%. Rents rise by 0.3% on average. Table 3 compares these changes with those from removing all admission restrictions. Covenants account for 1.8% of the resulting decline in non-White segregation and for 10.2% of the decline in White-ethnic segregation. The local effects are larger (Figure 5). In the 367 locations with covenants, the minority share rises by 13.3% relative to the baseline, and rents rise by 1.5%. In the other 3,570 locations, both changes are close to zero.

Figure 6: Counterfactual 2: Open Access



Note: This figure maps the change in the number of households of each group and the change in rents across developments when every group is admitted to every development. The bottom-right panel reports the average rent change, the change in segregation for each minority group against all other households, and the change in population and welfare for each group. The star marks the central business district.

In the second counterfactual, we remove all neighborhood choice-set restrictions. Removing selective admissions increases demand, and rents rise by 9.4%. Despite the higher rents, the minority share of households in the metro area rises from 7.6% to 9.3%. The number of non-White households rises by 39%, and the number of White-ethnic households by

Table 4: Open Access: Admission and Rent Channels

		Segregation			Welfare (CV)		
	Prices	Non-White	White-ethnic	Minority households	Non-minority	White-ethnic	Non-White
(1) Open admission, baseline prices	+0.0%	+0.5%	−0.7%	+33.3%	+0.0%	+19.5%	+29.0%
(2) Prices clear	+9.4%	−16.9%	−10.1%	−8.5%	−1.9%	−5.9%	−12.0%
Net effect	+9.4%	−16.4%	−10.8%	+22.0%	−1.9%	+12.4%	+13.5%

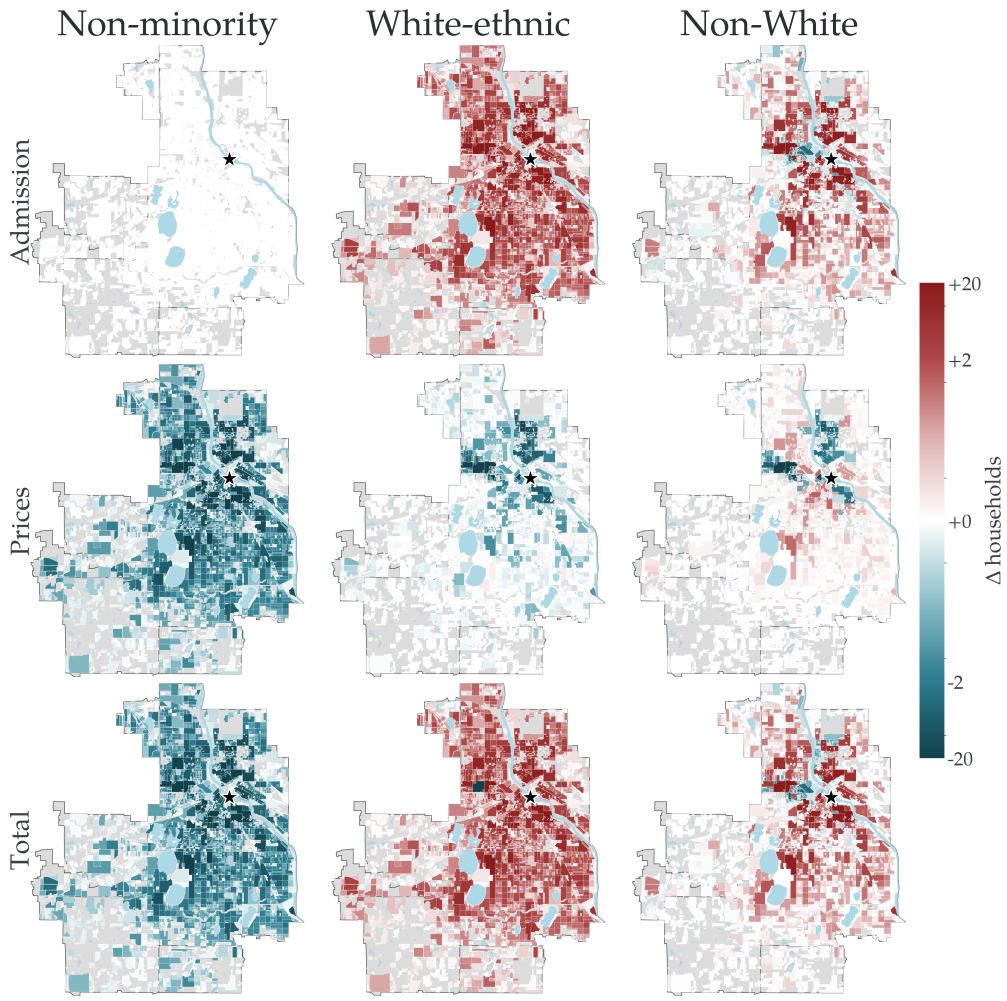
Note: This table decomposes the open-access counterfactual into two steps. Step (1) admits every group to every development at baseline rents. This state is off equilibrium. Step (2) lets rents clear each location. Segregation is the adjusted dissimilarity index at the ED level, for each group against all other households. Changes in segregation add across steps. Changes in households and welfare compound across steps. Welfare is the compensating variation as a percentage of each group's baseline rent.

19% (Table 2). These changes indicate that the choice-set restrictions disproportionately limited the entry of minority households into the metro area. Segregation falls by 16% for non-White households and by 11% for White-ethnic households. Figure 6 maps these changes across developments.

Table 4 separates the admission and rent channels. We first open admission at baseline rents. We then let rents clear the market. Opening admission alone raises the number of minority households in the metro area by 33%. It leaves segregation almost unchanged. Rents then rise most where minority demand increases, which moves some minority households out of those places. This rent step lowers the number of minority households by 8.5% and non-White segregation by 17%. Open access raises welfare by 13.5% of baseline rent for non-White households and by 12.4% for White-ethnic households. Non-minority households lose 1.9%, because they face higher rents without any change in access. Figure 7 maps the two channels.

In the third scenario, we eliminate households' preferences over neighborhood demographic composition in addition to all selective admissions. Rents fall by 1.6% on average, but rent changes are very dispersed. The standard deviation of the change in log rent is 0.975, against 0.068 under open access. We therefore read the rent results of this scenario with caution. The number of minority households in the metro area rises by 2.5%, less than the 22% increase under open access alone. Segregation falls by 66% for non-White

Figure 7: Open Access: Change in Households by Channel

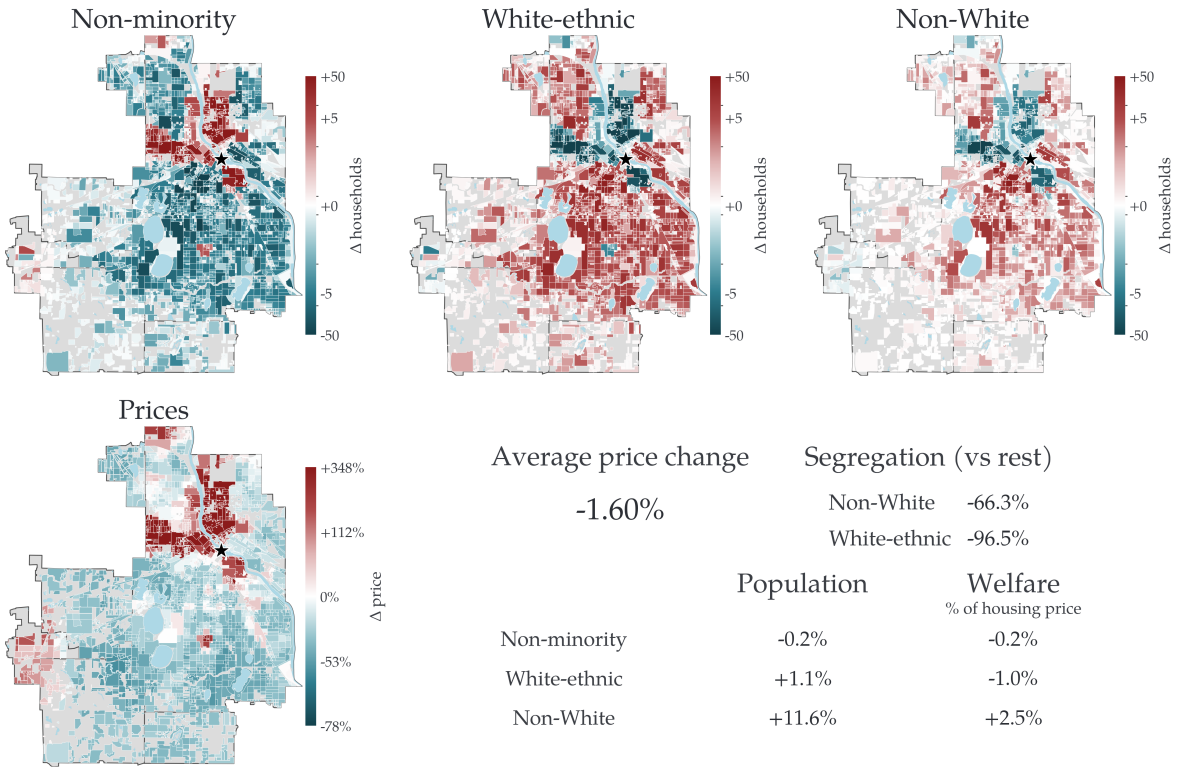


Note: This figure maps the change in the number of households of each group under open access. The top row opens admission at baseline rents. The middle row shows the additional change when rents clear the market. The bottom row shows the total. The star marks the central business district.

households and by 97% for White-ethnic households (Figure 8).

These findings suggest that preferences over neighborhood demographic composition are a more important driver of segregation than the instrument of racial covenants or choice set restrictions. The removal of all choice set restrictions explains about one-quarter of the decline in non-White segregation achieved when both choice set restrictions and preferences over neighborhood demographic composition are eliminated: a decline of 16.4% against 66.3%. For White-ethnic households, the share is about one-tenth: a decline of 10.8% against 96.5%. Note that the estimated distaste for minority households in 1940

Figure 8: Counterfactual 3: Open Access and No Homophily



Note: This figure maps the change in the number of households of each group and the change in rents across developments when every group is admitted to every development and preferences over neighborhood demographic composition are removed. The bottom-right panel reports the average rent change, the change in segregation for each minority group against all other households, and the change in population and welfare for each group. The star marks the central business district.

could be shaped by past policy or people's worldviews, independent of those factors. We are unable to separate these influences. Additionally, these preferences reflect not only prejudice but also homophily from cultural and linguistic similarities. Again, we cannot differentiate between these two aspects.

6. Conclusion

In this paper, we study how both household preferences and restrictions on households' choice sets—whether *implicit* or policy-driven, *explicit* discrimination—affect the observed prices and sorting of households across space. To do this, we have created a novel dataset that links households from the 1940 individual-count US census data to real estate developments in the 1940 Minneapolis metro area, and to the racial covenants and street grid

of that time. We propose a two-sided version of the traditional discrete-choice models of residential choice, where we model restrictions arising from both *implicit* and *explicit* discrimination as selective admissions of households on the supply side. We use our newly collected data to estimate our model, finding that minority households had lower admission probabilities into developments unlike the White non-minority households. We also find that ignoring restrictions on choice results in biased estimates for minority households, a result that aligns with our theoretical predictions. Our counterfactuals suggest that removing restrictions on households' neighborhood choice sets accounts for about one-quarter of the decline in non-White segregation achieved when both restrictions and preferences over neighborhood composition are removed. For White-ethnic households, it accounts for about one-tenth. The remaining decrease in segregation is attributed to the removal of distaste for minority neighbors.

References

Aaronson, Daniel, Daniel Hartley, and Bhashkar Mazumder, "The effects of the 1930s HOLC "redlining" maps," *American Economic Journal: Economic Policy*, 2021, *13* (4), 355–392.

Abaluck, Jason and Abi Adams-Prassl, "What do consumers consider before they choose? Identification from asymmetric demand responses," *The Quarterly Journal of Economics*, 2021, *136* (3), 1611–1663.

Agarwal, Nikhil and Paulo Somaini, "Demand Analysis under Latent Choice Constraints," *The Review of Economic Studies*, 10 2025, p. rdaf093.

Akbar, Prottoy A., Sijie Li Hickly, Allison Shertzer, and Randall P. Walsh, "Racial Segregation in Housing Markets and the Erosion of Black Wealth," *The Review of Economics and Statistics*, 2025, *107* (1), 42–54.

Albright, Alex, Jeremy A. Cook, James J. Feigenbaum, Laura Kincaide, Jason Long, and Nathan Nunn, "After the Burning: The Economic Effects of the 1921 Tulsa Race Massacre," Working Paper 28985, National Bureau of Economic Research 2021.

Almagro, Milena, Eric Chyn, and Bryan A. Stuart, "The Welfare Impacts of Urban Renewal and Forced Relocation: Evidence from Chicago's HOPE VI Demolitions," 2026. Working paper.

Anderson, Philip J. and Dag Blanck, eds, *Swedes in the Twin Cities: Immigrant Life and Minnesota's Urban Frontier*, St. Paul, MN: Minnesota Historical Society Press, 2001.

Bagagli, Sara, "The (Express)Way to Segregation: Evidence from Chicago," 2025. Working paper.

Bayer, Patrick, Fernando Ferreira, and Robert McMillan, "A unified framework for measuring preferences for schools and neighborhoods," *Journal of political economy*, 2007, *115* (4), 588–638.

_ , Marcus D. Casey, W. Ben McCartney, John Orellana-Li, and Calvin S. Zhang, "Distinguishing Causes of Neighborhood Racial Change: A Nearest-Neighbor Design," *American Economic Review*, 2025, *115* (11), 3999–4039.

Bergman, Klas, *Scandinavians in the State House: How Nordic Immigrants Shaped Minnesota Politics*, St. Paul, MN: Minnesota Historical Society Press, 2017.

Bergman, Peter, Raj Chetty, Stefanie DeLuca, Nathaniel Hendren, Lawrence F Katz, and Christopher Palmer, "Creating moves to opportunity: Experimental evidence on barriers to neighborhood choice," *American Economic Review*, 2024, *114* (5), 1281–1337.

Berry, Steven, James Levinsohn, and Ariel Pakes, "Automobile Prices in Market Equilibrium," *Econometrica*, 1995, *63* (4), 841–890.

Boustan, Leah Platt, "Was postwar suburbanization "white flight"? Evidence from the black migration," *The Quarterly Journal of Economics*, 2010, *125* (1), 417–443.

Brigham, Carl C., *A Study of American Intelligence*, Princeton, NJ: Princeton University Press, 1923.

Caetano, Gregorio and Vikram Maheshri, "A Unified Empirical Framework to Study Neighborhood Segregation," *Quantitative Economics*, 2025, *16*, 1023–1057.

Card, David, "Immigrant inflows, native outflows, and the local labor market impacts of higher immigration," *Journal of Labor Economics*, 2001, *19* (1), 22–64.

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

Carrington, William J. and Kenneth R. Troske, "On measuring segregation in samples with small units," *Journal of Business & Economic Statistics*, 1997, *15* (4), 402–409.

Chetty, Raj, Nathaniel Hendren, and Lawrence F Katz, "The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment," *American Economic Review*, 2016, *106* (4), 855–902.

Christensen, Peter and Christopher Timmins, "Sorting or steering: The effects of housing discrimination on neighborhood choice," *Journal of Political Economy*, 2022, *130* (8), 2110–2163.

_ and _ , "The damages and distortions from discrimination in the rental housing market," *The Quarterly Journal of Economics*, 2023, *138* (4), 2505–2557.

Chyn, Eric and Lawrence F Katz, "Neighborhoods matter: Assessing the evidence for place effects,"

Journal of Economic Perspectives, 2021, *35* (4), 197–222.

Collins, William J and Robert A Margo, "Residential segregation and socioeconomic outcomes: When did ghettos go bad?," *Economics Letters*, 2000, *69* (2), 239–243.

Conlon, Christopher T and Julie Holland Mortimer, "Demand estimation under incomplete product availability," *American Economic Journal: Microeconomics*, 2013, *5* (4), 1–30.

Cook, Cody, Pearl Z. Li, and Ariel J. Binder, "Where to Build Affordable Housing? Evaluating the Tradeoffs of Location," Working Paper CES-WP-23-62, U.S. Census Bureau, Center for Economic Studies 2023.

Couture, Victor, Cécile Gaubert, Jessie Handbury, and Erik Hurst, "Income Growth and the Distributional Effects of Urban Spatial Sorting," *The Review of Economic Studies*, 2024, *91* (2), 858–898.

Crawford, Gregory S, Rachel Griffith, and Alessandro Iaria, "A survey of preference estimation with unobserved choice set heterogeneity," *Journal of Econometrics*, 2021, *222* (1), 4–43.

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

Dahl, Gordon B., Andreas Ravndal Kostøl, and Magne Mogstad, "Family Welfare Cultures," *Quarterly Journal of Economics*, 2014, *129* (4), 1711–1752.

Diamond, Rebecca, "The determinants and welfare implications of US workers' diverging location choices by skill: 1980–2000," *American economic review*, 2016, *106* (3), 479–524.

Dobbie, Will, Jacob Goldin, and Crystal S. Yang, "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges," *American Economic Review*, 2018, *108* (2), 201–240.

Dubé, Jean-Pierre, Ali Hortacsu, and Joonhwi Joo, "Random-Coefficients Logit Demand Estimation with Zero-Valued Market Shares," *Marketing Science*, 2021, *40* (4), 637–660.

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

Fischel, William A., *The Homevoter Hypothesis: How Home Values Influence Local Government Taxation, School Finance, and Land-Use Policies*, Cambridge, MA: Harvard University Press, 2001.

Fishback, Price V., Jessica LaVoice, Allison Shertzer, and Randall P. Walsh, "The HOLC Maps: How Race and Poverty Influenced Real Estate Professionals' Evaluation of Lending Risk in the 1930s," *The Journal of Economic History*, 2023, *83* (4), 1019–1056.

Gallagher, Ryan, Allison Shertzer, and Tate Twinam, "The Long-Run Impacts of Zoning the Suburbs," *Working Paper*, 2024.

Gandhi, Ashvin, "Picking Your Patients: Selective Admissions in the Nursing Home Industry," 2026. Conditionally accepted at the American Economic Review.

Gaynor, Martin, Carol Propper, and Stephan Seiler, "Free to choose? Reform, choice, and consideration sets in the English National Health Service," *American Economic Review*, 2016, *106* (11), 3521–3557.

Goeree, Michelle Sovinsky, "Limited Information and Advertising in the U.S. Personal Computer Industry," *Econometrica*, September 2008, *76* (5), 1017–1074.

Gold, Daniel E, Lu Han, and Christopher Timmins, "Hidden Biases: Selective Advertising in the Rental Housing Market," *Manuscript*, 2025.

Goolsbee, Austan and Amil Petrin, "The consumer gains from direct broadcast satellites and the competition with cable TV," *Econometrica*, 2004, *72* (2), 351–381.

Honka, Elisabeth, Ali Hortaçsu, and Matthijs Wildenbeest, "Empirical Search and Consideration Sets," in "Handbook of the Economics of Marketing," Vol. 1, Elsevier, 2019, pp. 193–257.

Hoyt, Homer, *One Hundred Years of Land Values in Chicago: The Relationship of the Growth of Chicago to the Rise in Its Land Values, 1830–1933*, Chicago: University of Chicago Press, 1933.

Hynsjö, Disa M and Luca Perdoni, "The effects of federal "redlining" maps: a novel estimation strategy," in "Labor/Public Economics Workshop. Yale University Department of Economics" 2022.

Jr, Samuel L Myers, "Racial discrimination in housing markets: Accounting for credit risk," *Social Science Quarterly*, 1995, pp. 543–561.

Katz, Arnold J, "Imputing Rents to Owner-Occupied Housing by Directly Modeling Their Distribution," *WP2017-7, BEA Working Paper*, 2017.

Kling, Jeffrey R., "Incarceration Length, Employment, and Earnings," *American Economic Review*, 2006, *96* (3), 863–876.

Li, Nicholas Y, "Racial Sorting, Restricted Choices, and the Origins of Residential Segregation in US Cities," *George Washington University*, 2023.

Logan, Trevon D and John M Parman, "Segregation and homeownership in the early twentieth century," *American Economic Review*, 2017, *107* (5), 410–14.

Massey, Douglas S, "American apartheid: Segregation and the making of the underclass," *American journal of sociology*, 1990, *96* (2), 329–357.

Ouazad, Amine and Romain Rancière, "City equilibrium with borrowing constraints: Structural estimation and general equilibrium effects," *International Economic Review*, 2019, *60* (2), 721–749.

Potter, Constance, "The 1940 Census Revisited," *Prologue*, 2012, *44* (4).

Rothstein, Richard, *The Color of Law: A Forgotten History of How Our Government Segregated America*, illustrated edition ed., New York ; London: Liveright, May 2017.

Saiz, Albert, "The geographic determinants of housing supply," *The Quarterly Journal of Economics*, 2010, *125* (3), 1253–1296.

Schelling, Thomas C, "Models of segregation," *The American Economic Review*, 1969, *59* (2).

Shertzer, Allison and Randall P Walsh, "Racial sorting and the emergence of segregation in American cities," *Review of Economics and Statistics*, 2019, *101* (3), 415–427.

_ , Tate Twinam, and Randall P. Walsh, "Race, Ethnicity, and Discriminatory Zoning," *American Economic Journal: Applied Economics*, July 2016, *8* (3), 217–46.

Song, Jaehee, "The Effects of Residential Zoning in U.S. Housing Markets," *Journal of Urban Economics*, 2025, *149*, 103784.

Sood, Aradhya and Kevin Ehrman-Solberg, "The Long Shadow of Housing Discrimination: Evidence from Racial Covenants," *The Review of Economic Studies*, 2026.

Tabellini, Marco, "Gifts of the immigrants, woes of the natives: Lessons from the age of mass migration," *The Review of Economic Studies*, 2020, *87* (1), 454–486.

Train, Kenneth E, *Discrete choice methods with simulation*, Cambridge university press, 2009.

Trounstine, Jessica, *Segregation by design: Local politics and inequality in American cities*, Cambridge University Press, 2018.

Waldinger, Daniel, "Targeting in-kind transfers through market design: A revealed preference analysis of public housing allocation," *American Economic Review*, 2021, *111* (8), 2660–2696.

Weiwu, Laura, "Unequal Access: Racial Segregation and the Distributional Impacts of Interstate Highways in Cities," 2025. Revision requested at the American Economic Review.

Choices or Constraints? Disentangling Preferences from Restrictions in Housing Markets

by Milena Almagro and Aradhya Sood

ONLINE APPENDIX

A. Data Appendix

A.1 1940 Individual Count US Census Data

We use the 1940 individual count census data for the state of Minnesota from Integrated Public Use Microdata Series (IPUMS). Using the county identifier, we arrive at 570,044 individual records in Hennepin County in 1940. Aggregating up to the head of household level, there are 179,923 households across the whole of Hennepin County, which we restrict to 164,674 households when we focus only on the six municipalities in the sample of the Minneapolis metro area.

A.2 Housing Tenure and Monthly Rental Value Data and Imputation

Table A.1: Housing Tenure Shares by Minority Status

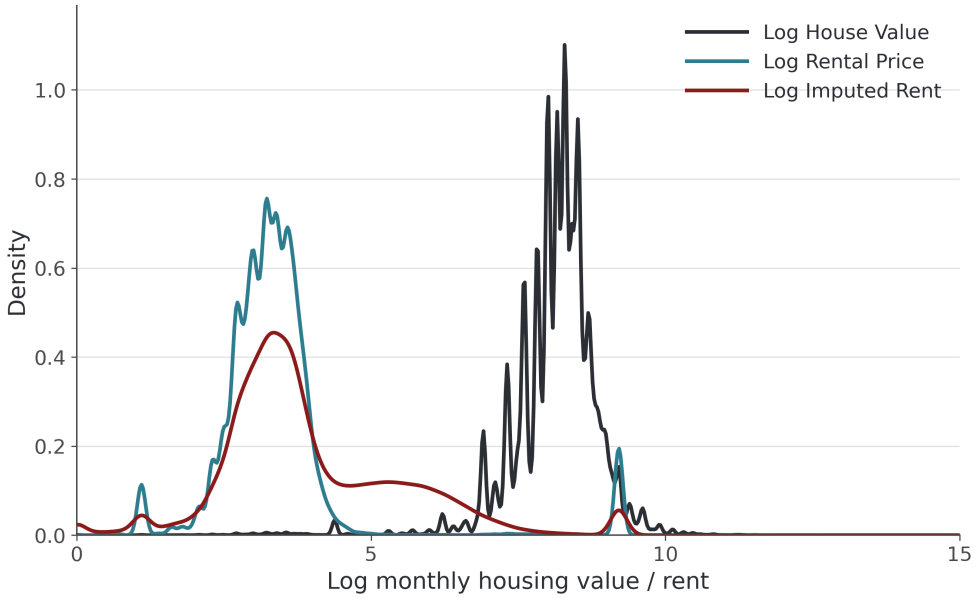
Group	Owner-Occupied	Renter-Occupied	Group Quarters
Non-Minority	0.39	0.54	0.07
Minority (Combined)	0.48	0.47	0.05
Minority (Non-White)	0.19	0.73	0.07
Minority (White Ethnic)	0.52	0.43	0.05

Note: This table shows housing tenure shares for owner-occupied, renter-occupied, and group quarter-occupied homes for non-minority and minority households. We also break down the housing tenure shares within the two minority household subgroups: non-White and White ethnic groups.

Table A.1 provides the housing tenure shares for owner-occupied, renter-occupied, and group quarter-occupied homes for non-minority and minority households. We also break down the housing tenure shares within the two minority household subgroups: non-White and White ethnic groups. As can be seen from the table, there is variation in home ownership rates across different demographic groups.

As noted in Section 2.2, we only observe monthly rental prices and housing values for renter-occupied and owner-occupied households, respectively. To enable a comparison

Figure A.1: House Value, Rental Price, and Imputed Rental Value Distributions

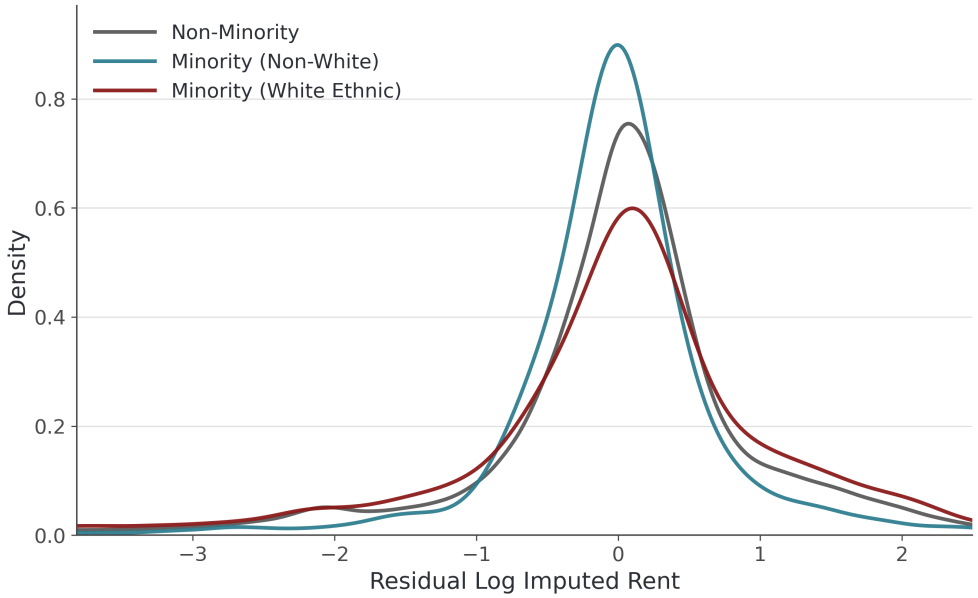


Note: This figure plots the log of house values for owner-occupied housing (in black) and the log of monthly rents for renter-occupied housing (in green) for the households in the analysis sample. In purple, it plots the combined log monthly rents for renter-occupied housing and the imputed monthly owner cost of housing for owner-occupied housing.

between monthly rents and housing values, we impute monthly rental values for owner-occupied housing units. Following Katz (2017), we use a gamma distribution with shape parameter 0.8 and scale parameter 1 to impute annual rental values, which we divide by 12 to obtain monthly rental values. The imputed rental value of owner-occupied housing, combined with rental price data, yields the mean development-level monthly rental value for 98.2% of the developments in our sample. Figure A.1 plots the log of house values for owner-occupied housing (in black) and the log of monthly rents for renter-occupied housing (in green) for the households in the analysis sample. In purple, it plots the combined log monthly rents for renter-occupied housing and the imputed monthly owner cost of housing for owner-occupied housing.

Figure A.2 plots the residuals from a regression of the log of monthly rental values on development fixed effects interacted with housing tenure fixed effects, capturing the variation in rental value within developments. As noted in Section 3.1.2, we do not account for the differences in rental values paid by various demographic groups. This is because the figure illustrates that the rents paid by non-White minority households

Figure A.2: Distribution of Monthly Rental Value Residuals



Note: This figure plots the residuals from a regression of the log of monthly rental values on development fixed effects interacted with housing tenure fixed effects.

are only slightly lower within developments. In contrast, the rents paid by White ethnic minority households exhibit a distribution similar to that of non-minority households.

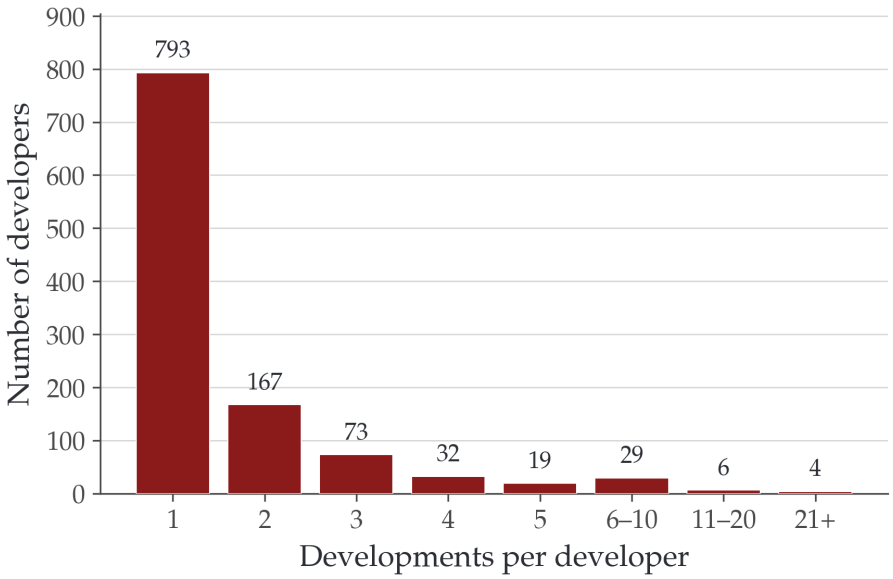
A.3 Data on Developers

We identify developers from historical development maps obtained from Hennepin County's land records (see Figure F.4 for an example). We record up to five individual developers or platters and any developer or developer firm and combine manual transcriptions of these maps with AI-assisted visual readings, using two independent AI agent readings and a third AI agent reader to adjudicate disagreements, supplemented by manual review. For developments absent from current parcel records, we locate historical developments using development names and historical maps. Racial covenant grantor records from Mapping Prejudice also supply developer companies for three additional developments. We retain developer identities where the associated development subdivision year is observed and no later than 1940. In total, we are able to identify developers for 2,008 developments out of 3,158 developments.

We then link developers across developments by standardizing names and matching spelling variants, initials and firm names alongside manual review. Each original de-

velopment is assigned to the recorded firm, where available, and otherwise to the first recorded developer. For developments combining several original developments, we use the most frequent linked identity, giving priority to a firm in a tie. Figure A.3 plots the distribution of developers across the number of developments. Most identified developers (793 of 1,123, or 71%) are linked to just one development, while a small number are linked to many developments.

Figure A.3: Distribution of Developments per Developer



Note: This figure reports the number of developments associated with each identified developer. Each developer is counted once. The sample contains 1,123 identified developers associated with 2,008 developments. Developments without an identified developer are excluded.

A.4 Data on Historic Real Estate Developments

To construct the geography of historic real estate developments, we begin with Hennepin County assessor parcel data and group parcels by development ID. This yields 5,843 present day development polygons in the Minneapolis metro area once we clean boundaries and split multipart geometries into contiguous pieces. We then add and revise geometries using digitised historical city atlases with development maps (see Figure F.3 for an example page from the city atlas) for places where current parcels are missing, destroyed, or inconsistent with historical development boundaries. After incorporating

these historical supplements and replacements, the map contains 6,561 contiguous polygons. This process also includes grouping polygons with identical development names that are within 500 meters of each other. The 6,561 contiguous polygons collapse to 3,160 unique developments that contain at least one household in the 1940 census.

We obtain platting dates from historical plat maps. For unplatted tracts identified by section, township and range, we supplement these dates with tract cards. We retain dates no later than 1940.

A.5 Data on Racial Covenants

The Mapping Prejudice Project searched every sales deed and development map from 1900 to 1960 in Hennepin County to identify those with racial covenants. There are 25,818 racial covenants at the lot level in Hennepin County, of which we identify 24,642 as unique covenant observations based on the following attributes: language used on covenants, development name, city, block, lot, grantor and grantee name, date of deed, and date of execution. Of those, there are 19,668 unique observations in the six municipalities to which we restrict our attention. We use the historical development-level dataset and Hennepin County assessor data to match racial covenants to developments. We match covenant information to parcels within each development to determine the share of covenants in each development. In addition, we also keep track of the date of the covenant execution.

A.6 Historic Street Grids

We use historic street network data to geocode addresses and developments. For the city of Minneapolis, our primary source is the Map USA Database from Brown University (Logan, 2017). To enhance this data, we also collect information from city atlas maps and Sanborn insurance maps for Minneapolis, as well as for five other municipalities: St. Louis Park, Edina, Richfield, Hopkins, and Robbinsdale, in and around 1940. The street grid files have been collected, digitized, geocoded, and cleaned to extract details such as street names, types (e.g., street, avenue, boulevard), and directional prefixes or suffixes. Additionally, these files include interpolated address number ranges for accurate geocoding.

A.7 Data on Enumeration District and Municipality Boundaries

To define broader neighborhoods within a municipality, we digitized and geocoded the 1940 census enumeration district (ED) boundary files for Hennepin County. We performed extensive data cleaning to standardize the ED numbering system and align the boundaries with other geographic datasets. These boundary files enable us to map the census household data accurately within each ED. Additionally, we use municipality boundary shapefiles to fill in any missing municipalities in household addresses. Because there is a direct association between an ED identifier and a municipality, there is no risk of mistakenly assigning a household to the wrong municipality.

For the ED-level shift-share instruments, we express 1930 population counts on 1940 ED boundaries. We intersect the two sets of ED polygons and allocate each 1930 ED's counts across 1940 EDs in proportion to the area of overlap. This effectively assumes that each group's population is uniformly distributed within a 1930 ED. The method follows the same area weighting principle used by Shertzer and Walsh (2019) and Lee and Lin (2018).

A.8 Housing Characteristics and Other Geographic Data

The baseline lot size and build year data, as mentioned in Section 2.2, are obtained from the Hennepin County Assessor's Office. If either of these data points is missing, we supplement it with Zillow data. Please note that while Zillow's data was freely available under an academic license in 2019, it is no longer offered for free.

For Minneapolis developments without a usable building year, we supplement these sources with digitised historical building permit cards from Hennepin County Library.[1]

We match these records by address or addition name, using streets, blocks and lots from the 1940 city atlas to resolve additional matches. We infer building completion as the year of the first dated A or B construction permit plus one, and retain recovered building years no later than 1940.

The slope data come from the Natural Resources Conservation Service. We use the mean elevation and slope at the development level. We digitize and geocode a 1899 map of

[1]The digitised records are available here.

Hennepin County to highlight wetlands and semi-submerged areas, and use this to create the share of development with wetlands measure. To assess proximity to commercial and industrial zones, we create a shapefile based on the 1940 Minneapolis city atlas (plate 84B), reflecting the 1923 zoning ordinance for the city of Minneapolis. We calculate the Euclidean distance from each development's edge to heavy industrial and commercial areas. We acknowledge that, while we don't consider commercial and industrial zones in five other suburban municipalities, Minneapolis was the metro area's primary commercial and industrial hub. We also assume that activities occur within their designated zones, as land use is generally adhered to in the US, including in the mid-20th century (Shertzer et al., 2016). We also digitize the 1913 streetcar map of the Minneapolis metro area and calculate the Euclidean distance from the edge of a development to the nearest point on the streetcar path. Lastly, we also calculate the Euclidean distance from each development to its nearest lake and to the central business district (CBD) of the metro area, St. Anthony Falls, where the commercial and industrial hub of the area resided in the mid-20th century.

A.9 Imputation of Missing Household Addresses

To enable household-development level analysis, we cleaned, geocoded, and matched a total of 164,674 households from six municipalities included in our sample. The initial step involved using raw addresses obtained from the census, which we meticulously cleaned to extract house numbers, directional prefixes, street names, and street types. Before any processing, we found that 95.6% of the addresses contained a usable street name, 78.1% included a house number, and 75.1% had both components (see Table A.2, Panel A). It is important to note that every household in the census is assigned an enumeration district (ED) ID, which also provides a link to the municipality. Following the methodology outlined by Logan and Parman (2017) as well as using string parsing, pattern matching, and fuzzy text matching with reference street-name files, we filled missing components using adjacent records. As a result of this multi-step cleaning and imputation process, coverage improved to 96.9% for street names, 98.3% for house numbers, and 96.1% for complete addresses (Table A.2 Panel B).

The cleaned addresses were then geolocated by matching them to Hennepin County parcel and street grid files. This matching process utilized the ED, house number, street

Table A.2: House Address Imputation

Address Information	Count	%
Panel A: Raw		
Street Name	157,452	95.61
House Number	128,596	78.09
Full Addresses	123,721	75.13
Total	164,674	100.00
Panel B: After Imputation		
Street Name	159,536	96.88
House Number	161,911	98.32
Full Addresses	158,265	96.11
Total	164,674	100.00

Note: Panels A and B in the table display the counts and percentages (%) of households with address information in the raw 1940 census data and after the imputation process, respectively, in the six municipalities in the sample. "Street Name" indicates the availability of the street name, "House Number" indicates the presence of a house number, and "Full Addresses" indicates households with both street name and house number.

name, street type, and directional information, resulting in 40.6% of addresses being successfully matched. The remaining addresses were matched manually using historical development shapefiles and the street grid. Of the 164,674 households, 96.1% were matched exactly to a development, while the remaining households were assigned randomly to a development within the same ED.

A.10 Combining Datasets

The final step in the data construction process involved spatially matching the various datasets together. The output is a household-development level dataset for the Minneapolis metro area. Table A.3 provides summary statistics for the analysis sample at the household, development, and enumeration district levels. P_{25}, P_{50}, and P_{75} refer to the 25th, 50th, and 75th percentile, respectively.

A.11 Development–Enumeration District Correspondence

As highlighted in Section 2.3, the developments do not perfectly coincide with the enumeration districts (EDs) in the Census. Around 500 developments contain households

belonging to different EDs, usually 2 different EDs. We, therefore, perform the analysis at the unit of the intersection of developments and EDs, of which there are around 4000. In this section, we show that our estimates are unbiased and efficient if we do so. In simple household-level model,

$$y_i = \beta D_{j(i)} + \gamma X_{j(i)} + \lambda_{e(i)} + u_i,$$

i indexes households, $j(i)$ denotes the development containing household i, and $e(i)$ denotes the ED containing household i. The treatment D_j and covariates X_j vary at the development level, while λ_e is an ED fixed effect.

As long as treatment and covariates vary only at the development level, the model can be aggregated without loss of information to the intersection of developments and EDs. Let N_{je} denote the number of households in the intersection of development j and ED e, and define the mean outcome at this level as:

$$\bar{y}_{je} \equiv \frac{1}{N_{je}} \sum_{i \in j \cap e} y_i.$$

Taking the mean of the household equation over all households in $j \cap e$ yields

$$\bar{y}_{je} = \beta D_j + \gamma X_j + \lambda_e + \bar{u}_{je},$$

where

$$\bar{u}_{je} = \frac{1}{N_{je}} \sum_{i \in j \cap e} u_i.$$

This expression shows that the household model aggregates exactly to the development–ED intersection level. The right-hand side variables remain unchanged because they are constant within each cell: D_j and X_j vary only by development, while λ_e varies only by ED.

To estimate this equation we can run a weighted regression at the (j, e) level with weights N_{je}. This aggregation is lossless in the sense that the weighted least squares objective function is identical to the household-level OLS objective. Specifically, household

OLS minimizes

$$\sum_{i} \left(y_i - \beta D_{j(i)} - \gamma X_{j(i)} - \lambda_{e(i)} \right)^2 .$$

Grouping households by (j, e) cells, this objective can be written as

$$\sum_{j,e} \sum_{i \in j \cap e} \left(y_i - \beta D_j - \gamma X_j - \lambda_e \right)^2 .$$

Using the standard decomposition of squared deviations,

$$\sum_{i \in j \cap e} (y_i - \mu)^2 = \sum_{i \in j \cap e} (y_i - \bar{y}_{je})^2 + N_{je} (\bar{y}_{je} - \mu)^2,$$

the first term does not depend on the parameters $(\beta, \gamma, \lambda_e)$. The second term implies that minimizing the household objective is equivalent to minimizing

$$\sum_{j,e} N_{je} \left(\bar{y}_{je} - \beta D_j - \gamma X_j - \lambda_e \right)^2 .$$

This is exactly the weighted least squares problem obtained by estimating the aggregated equation at the (j, e) level with weights N_{je}. Therefore, the coefficient estimates from the two regressions coincide, implying that aggregation from households to development–ED cells is lossless.

This equivalence also implies that the estimate of β obtained from the aggregated regression is unbiased and efficient under the same conditions as the household regression. Unbiasedness follows because

$$E[\bar{u}_{je} \mid D_j, X_j, \lambda_e] = 0$$

whenever the household disturbances satisfy

$$E[u_i \mid D_{j(i)}, X_{j(i)}, \lambda_{e(i)}] = 0.$$

Efficiency follows because weighted least squares with weights N_{je} reproduces the same

normal equations as household OLS, and therefore, uses all available information in the sample. Operationally, each development–ED intersection is treated as an observation weighted by the number of households in that cell, ensuring that estimation exactly replicates the household-level regression with ED fixed effects.

Table A.3: Statistics at the Household, Development, and Enumeration District Level

Variables	P_{25}	P_{50}	P_{75}	Mean	SD	N
Panel A: Enumeration District Level						
Number of Developments	4.00	7.00	10.00	8.42	7.48	468
Size (km^2)	1.71	2.40	4.54	6.18	22.01	468
Panel B: Development Level						
Number of Households	4.00	14.00	48.00	52.15	128.70	3,158
Size (km^2)	0.02	0.11	0.39	0.92	8.44	3,158
Panel C: Development–ED Level						
Number of Households	4.00	15.00	47.00	41.78	72.12	3,941
Size (km^2)	0.02	0.12	0.37	0.73	7.41	3,941
Minority Share	0.00	0.01	0.07	0.08	0.17	3,941
Share Covenanted	0.00	0.00	0.00	0.05	0.21	3,941
Mean Build Year	1908	1917	1926	1917	11	3,941
Mean Log Lot Size	8.63	8.90	9.64	9.36	1.16	3,941
Distance to Commercial Area (m)	1	69	345	510	1137	3,941
Distance to CBD (m)	3196	4958	8001	5744	3279	3,941
Distance to Heavy Industry (m)	144	525	1367	1052	1401	3,941
Distance to Nearest Lake (m)	468	1044	1769	1235	957	3,941
Distance to Streetcar (m)	121	286	760	684	985	3,941
Percent Wetlands	0.00	0.00	0.00	3.59	14.90	3,941
House Rental Values (USD Monthly)	43.62	110.20	208.84	182.55	319.93	3,937
Urbanization Index	0.00	0.26	0.71	0.36	0.36	3,941
Log Slope	0.71	1.06	1.20	1.01	0.60	3,941
Shift-Share Instrument	0.0004	0.0005	0.0006	0.0005	0.0003	3,941
Panel D: Household Level						
No. of Household Members	2.00	3.00	4.00	3.17	2.00	164,674
No. of Working Household Members	2.00	2.00	2.00	1.79	0.41	164,674
Home Ownership Dummy	0.00	0.00	1.00	0.40	0.49	164,674
Married Dummy	0.00	1.00	1.00	0.71	0.45	164,674

Note: This table provides summary statistics at household level, development level, and enumeration district level. P_{25}, P_{50}, and P_{75} refer to the 25th, 50th, and 75th percentile, respectively. Size is measured in square kilometers (km^2), while distance is measured in meters (m). CBD stands for cental business district.

References

Katz, Arnold J., "Imputing Rents to Owner-Occupied Housing by Directly Modelling Their Distribution," Working Paper WP2017-7, Bureau of Economic Analysis August 2017.

Lee, Sanghoon and Jeffrey Lin, "Natural Amenities, Neighbourhood Dynamics, and Persistence in the Spatial Distribution of Income," *The Review of Economic Studies*, 2018, *85* (1), 663–694.

Logan, John, "Map USA: A Human-Mapping Project (1940s–2010)," Brown University 2017.

Logan, Trevon D. and John M. Parman, "Segregation and Homeownership in the Early Twentieth Century," *American Economic Review*, 2017, *107* (5), 410–414.

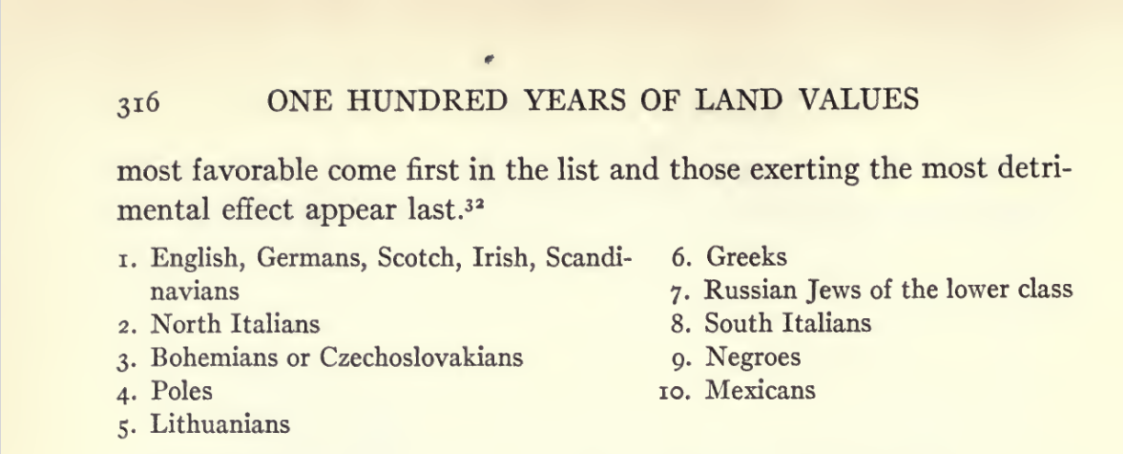
Shertzer, Allison and Randall P. Walsh, "Racial Sorting and the Emergence of Segregation in American Cities," *Review of Economics and Statistics*, 2019, *101* (3), 415–427.

_ , Tate Twinam, and Randall P. Walsh, "Race, Ethnicity, and Discriminatory Zoning," *American Economic Journal: Applied Economics*, 2016, *8* (3), 217–246.

B. Additional Empirical Evidence

As highlighted in Section 2.4, we group White immigrant households into two categories based on their listed birthplace from "most favorable" to "least favorable" following the ranking by Hoyt (1933). Figure B.1 shows these rankings. Below we describe which country we place in which of the two White immigrant categories:

Figure B.1: Homer Hoyt's Ranking of Households Based on Race and Ethnicity



316 ONE HUNDRED YEARS OF LAND VALUES

most favorable come first in the list and those exerting the most detrimental effect appear last.[32]

1. English, Germans, Scotch, Irish, Scandinavians
2. North Italians
3. Bohemians or Czechoslovakians
4. Poles
5. Lithuanians
6. Greeks
7. Russian Jews of the lower class
8. South Italians
9. Negroes
10. Mexicans

Note: This figure is from "One hundred years of land values in Chicago: The relationship of the growth of Chicago to the rise of its land values, 1830-1933" by Homer Hoyt (1933).

Northern & Western Europe: Denmark, Finland, Iceland, Norway, Sweden, the Netherlands, England, Scotland, Wales, the United Kingdom, and Ireland, as well as the British Commonwealth countries including Canada, Australia, New Zealand, Belgium, France, Liechtenstein, Luxembourg, Monaco, Switzerland, and Italy.

Eastern & Southern Europe and Other Regions: Albania, Portugal, Spain, Bulgaria, Hungary, Yugoslavia, Central Europe, Estonia, Latvia, Austria, Germany, Poland, Romania, Czechoslovakia, Other USSR/Russia, Turkey, Greece, Syria, Israel/Palestine, Lebanon, China, Japan, Korea, Indonesia, Philippines, India, Asia Minor, Asia (unspecified), Africa (unspecified), Puerto Rico, Mexico, Central America, Cuba, West Indies, South America, Unspecified.

Figure B.2: Brigham's Historical Eugenic Comparisons of Immigrant Origins

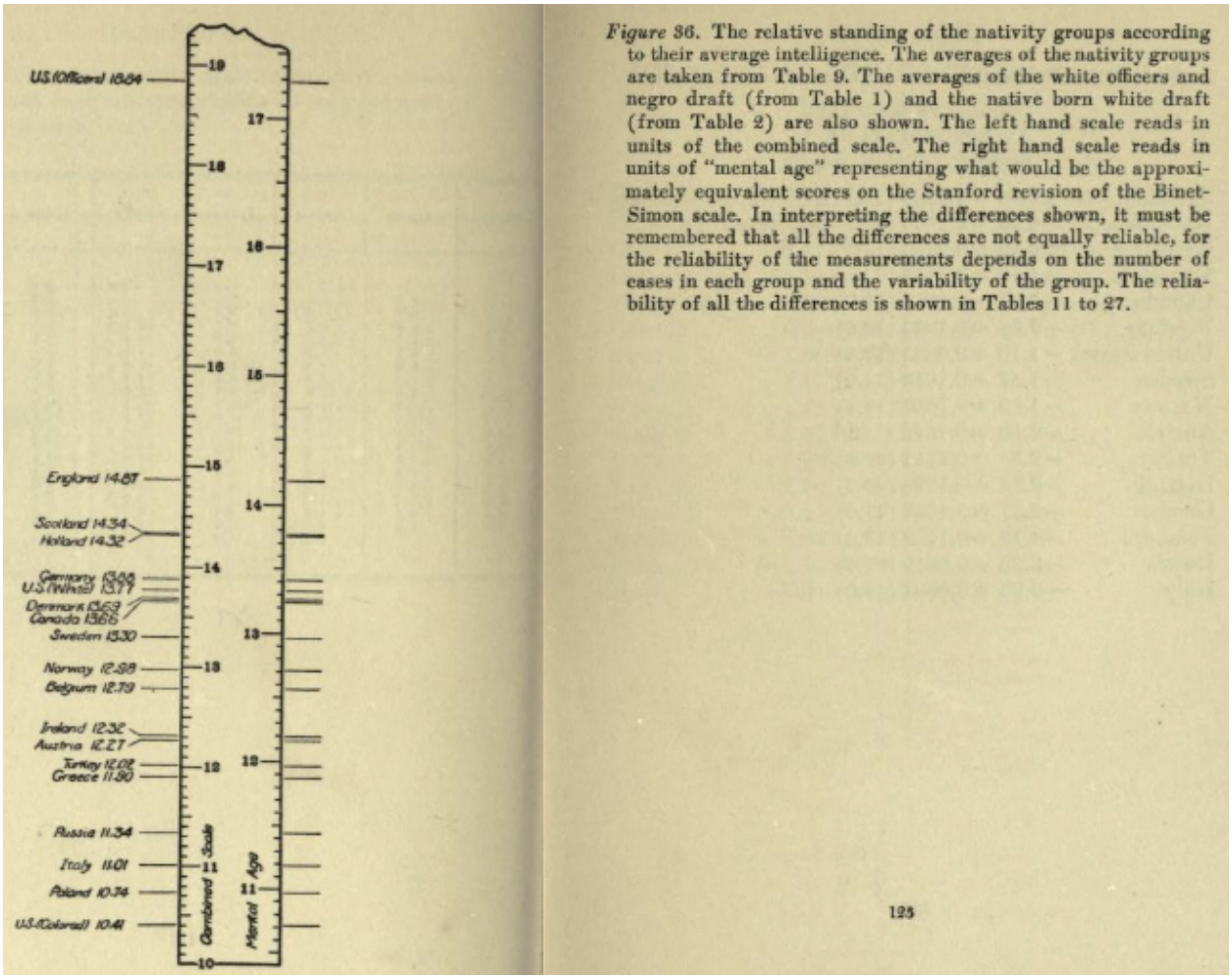


Figure 36. The relative standing of the nativity groups according to their average intelligence. The averages of the nativity groups are taken from Table 9. The averages of the white officers and negro draft (from Table 1) and the native born white draft (from Table 2) are also shown. The left hand scale reads in units of the combined scale. The right hand scale reads in units of "mental age" representing what would be the approximately equivalent scores on the Stanford revision of the Binet-Simon scale. In interpreting the differences shown, it must be remembered that all the differences are not equally reliable, for the reliability of the measurements depends on the number of cases in each group and the variability of the group. The reliability of all the differences is shown in Tables 11 to 27.

125

Notes: Reproduced from Brigham (1923, pp. 124–125). The left panel reproduces Figure 36, which ranks groups by country of birth and racial classification using Army test scores. The right panel contains Brigham's explanation of the figure, including the scales and qualifications concerning the reliability of group differences. We reproduce these pages as evidence of the contemporary eugenic racial and ethnic hierarchies.

B.1 Reduced-Form Model

To study the effects of racial covenants on neighborhood demographics, we estimate the following model, where i indicates a household and j indicates a development in 1940:

$$y_{ij} = \beta_0 + \beta_1 \text{Cov-Share}_{j(i)} + \beta_2 X_j + \eta_k(i) + \epsilon_{ij} \tag{B.1}$$

In this model, y is an indicator for a household's race or ethnicity in development j. Cov-Share$_{j(i)}$ is the share of the development j that had racial covenants.[2] X_j represents development level quality variables including the mean log of lot size, mean build year, the proportion with wetlands and the log of distance to each of a heavy industry zoned area, a commercially zoned area, the nearest lake, the streetcar, and the metro area's CBD.[3] $\eta_k(i)$ indicates the enumeration district k fixed effects within which household i is located.

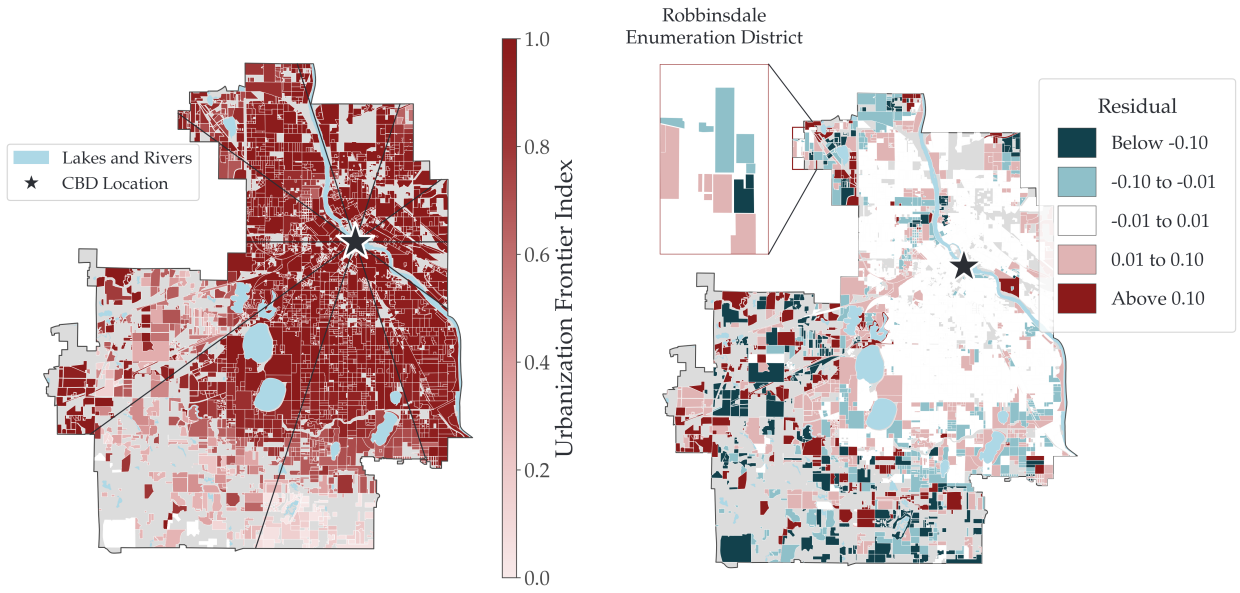
Given that the decision to add covenants is strategic, it is important to account for the non-random placement of racial covenants across space. Moreover, Sood and Ehrman-Solberg (2026) show that racial covenants were more prevalent in lower-middle and middle-income neighborhoods, serving as a marginal dissuasive factor in areas more likely to undergo demographic transitions in the future. While we control for enumeration district fixed effects, the placement of covenants may still be correlated with unobservable factors even within an enumeration district. To address this concern, we construct an instrument for the share of covenanted development.

To construct the instrument, we note three key facts: First, there is a clear radial outward expansion of new development away from the central business district (CBD), as shown in Figure 1 a-c. Second, racial covenants gained popularity nationwide between 1910 and 1940, as discussed in Section 2.1. Third, 94% of the racial covenants were added by developers on new housing and developments. Together, these imply that developments closer to the metro area's urbanizing frontier, or further from the CBD, were more likely to

[2] As illustrated in Appendix Figure B.4, usually, developments either have no covenants—of which there are many—or all houses within a development have covenants. However, there are a few developments that fall within the 1% to 99% range for covenant share. This variation is why the key independent variable is a share rather than a binary indicator.

[3] Racial covenant clauses may be correlated with unobserved other types of covenants or housing characteristics. Thus, controlling for the lot size helps control for building quality, while controlling for the building year controls for the development's "newness."

Figure B.3: Urbanization Frontier Index and Residual Patterns



Note: The left panel displays the urbanization frontier index across the ten radian sectors of 36° radiating from the central business district (CBD), which is displayed with a red diamond. The right panel shows the residuals from regressing the urbanization frontier index on the set of controls from Equation B.2 and enumeration district (ED) fixed effects. To highlight the spatial variation, we provide a zoomed-in image of an ED in Robbinsdale.

include racial covenants. Therefore, our instrument leverages the notion that developers between 1930 and 1940 were both likely to build along the 1930 urban frontier and also more inclined to use racial covenants.

We construct the instrument using the share of built-up area from previous periods in neighboring developments nearer to the city center. This instrument, *urbanization frontier index* denoted as f_j, captures the direction and momentum of urban expansion. To construct the index in 1940, we first divide the city from the CBD into r radian sectors of equal size. Within a sector r, we define $f_j \equiv \frac{\text{Built}_{r,1930}(z_j)}{\text{Lots}_{r,1930}(z_j)}$, where $\text{Built}_{r,1930}(z_j)$ is the number of houses built by 1930, and $\text{Lots}_{r,1930}(z_j)$ is the total number of buildable lots by 1930 within a given buffer zone z_j. Our preferred specification uses $r = 10$ and a buffer size of $z_j = 1000$ meters. Note that f_j ranges from 0 to 1. Figure B.3 (left panel) plots the urbanization frontier index as well as the radian sectors for all developments in the Minneapolis metro area for 1940, where the pre-period is 1930.[4]

The relevance condition is based on the radial outward expansion of construction seen

[4]The ten-year interval (1930-40) allows for sufficient time for housing construction to occur.

in Figure 1. As racial covenants became more widespread in the first half of the 20th century, our hypothesis posits that areas closer to this 1930 urbanized frontier were more likely to include them. In other words, our frontier index is a predictor of new housing construction during times when racial covenants are popular. Importantly, because we control for build year, the instrument does not simply capture new construction; rather, it predicts construction occurs during a decade when racial covenants were at the height of their popularity.

The exclusion restriction requires that, conditional on enumeration district fixed effects and development and location controls, the share of land developed by 1930 was as good as random. In this context, both land availability as well as idiosyncratic changes in the fortunes of developers and builders who had purchased land from farmers and other landowners provide the underlying source of quasi-random variation.[5] To see the variation that our identification strategy is capturing, we plot the residuals from regressing the urbanization frontier index instrument, f_j, on development and location controls as well as enumeration district fixed effects. Figure B.3 (right panel) plots the residuals for all developments, providing a closer view of the variation within a specific ED in Robbinsdale. As can be seen from the figure, there is random variation across developments, even within an ED. This variation shows that, conditional on the segment of the urban frontier, the exact location of real-estate development was random.

The first stage is given by the following model:

$$\text{Cov-Share}_{j(i)} = \alpha_0 + \alpha_1 f_j + \alpha_2 X_j + \eta_{k(i)} + \varepsilon_{j(i)}. \tag{B.2}$$

Observe that the instrument exploits temporal variation in the addition of covenants and ensures that development demographic differences within an enumeration district operate only through the addition of racial covenants at the frontier of the new development.

[5]Until the mid-1910s, the Minneapolis metro area concentrated commercial and residential development near the CBD. However, as the milling district declined in the early 1920s, development shifted outward (Danbom, 2003). This period was also difficult for U.S. farmers, who faced rising debt and falling commodity prices following World War I, leading to financial distress and farmland sales (Irwin, 2011). Using survey data from the 1960s and 1970s in neighboring Olmsted County, Minnesota, Pyle (1985) finds that the most common reason for farm sales for housing development was idiosyncratic personal shocks related to health and finances.

B.1.1 Racial Covenants and Sorting

Table B.1: Racial Covenants and Sorting in the Housing Market

	First Stage	Non-White	White Northern European Imm.	White Western European Imm.	White Eastern & Southern European Imm.	White Other Imm.
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: First Stage						
Urbanization Index	0.164***					
	(0.017)					
Panel B: OLS						
Share Covenanted		−0.006***	0.022*	−0.002	−0.001	−0.012**
		(0.001)	(0.011)	(0.002)	(0.001)	(0.005)
Panel C: IV 2SLS						
Share Covenanted		−0.100***	−0.051	0.048**	−0.028**	−0.004
		(0.034)	(0.149)	(0.023)	(0.013)	(0.083)
Location controls	✓	✓	✓	✓	✓	✓
ED FEs	✓	✓	✓	✓	✓	✓
Mean log lot size	✓	✓	✓	✓	✓	✓
Mean build year	✓	✓	✓	✓	✓	✓
Observations	164,673	164,673	164,673	164,673	164,673	164,673
Mean dependent variable	0.012	0.010	0.151	0.005	0.003	0.064
First-Stage F-Stat	89.9	89.9	89.9	89.9	89.9	89.9

Note: This table plot the first stage, OLS, and instrumental variable (IV 2SLS) parameter estimates from Equations B.1 and B.2. Non-White refers to households identified as Black, Asian, or Native American, regardless of their country of birth. White Northern European, Western European, Eastern and Southern European, and Other Immigrants (Imm.) denotes households identified as White and birthplace in these regions. Refer to Appendix B to see which countries are in each of these four categories. Panel A reports the first-stage parameter estimate of the urbanization frontier index (instrument) on share covenanted, where share covenanted is the fraction of lots in a development with racial covenants. Panels B and C show the OLS and IV 2SLS parameter estimates, respectively. Robust standard errors in parentheses. Location controls include log slope, wetlands share, and log distances to the central business district, heavy-industry zoning, commercial zoning, the nearest lake, and the streetcar. ED FEs are enumeration-district (ED) fixed effects. Mean log lot size and build year control for the development-level means of these variables. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.1 reports the parameter estimates α_1 and β_1 from Equations B.2 and B.1 with Panels B and C showing OLS and instrumental variable (IV 2SLS) estimates, respectively. The first-stage estimates in Panel A indicate that the instrument functions as intended; as

the urbanized built-up area increases, the likelihood of adding covenants in development j also rises. Specifically, moving from the minimum to the maximum value of the urbanization index (0 to 1) is associated with a 16.4 percentage point increase in the share of the development being covenanted. The first-stage F-Stat is 89.9.

In a fully covenanted development, an increase in Cov-Share$_{j(i)}$ from 0 to 1 results in a 10 percentage point decrease in the probability of a non-White household (column 2, Panel C).[6] This finding holds true after controlling for development characteristics, such as build year and lot size, as well as ED fixed effects that capture broader neighborhood factors like racial and immigrant concentration, housing quality, and local amenities at the ED level.

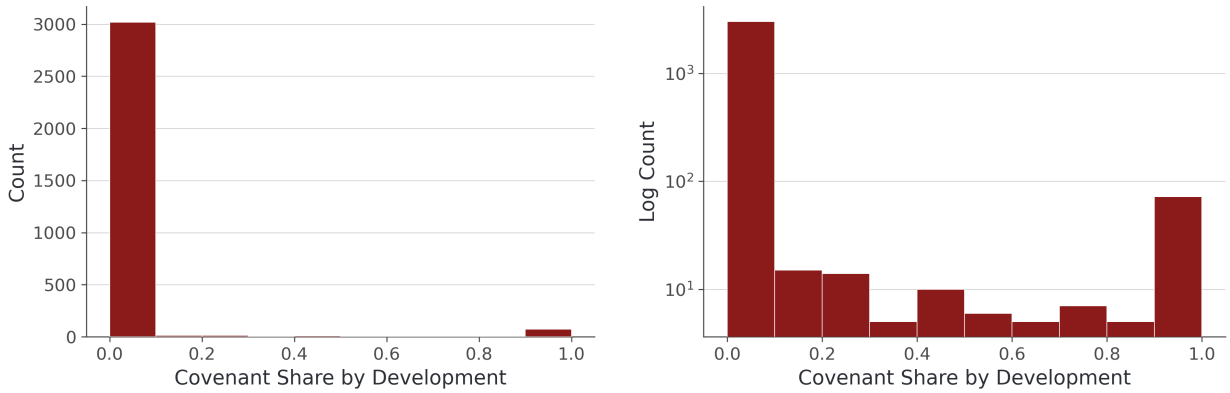
Using Hoyt's categories (Hoyt, 1933), our analysis shows that fully covenanted developments increased the likelihood of Western European White households by 4.8 percentage points (column 4, Panel C), indicating they were preferred residents and faced no discrimination. We did not find any statistically significant effects for White households from Northern Europe. In contrast, covenants reduced the likelihood of Eastern and Southern European households by 2.8 percentage points (column 5, Panel C), highlighting that these covenants were a tool for discrimination not only against racial minorities but also against specific White ethnic groups during the early-to-mid 20th century.

The IV estimates (Panel C) are larger in absolute terms than the OLS estimates (Panel B), which may reflect two sources of bias in the OLS estimates. First, unobserved neighborhood characteristics may be positively correlated with both the adoption of covenants and the presence of households that did not face neighborhood restrictions, biasing the OLS coefficients toward zero. Second, endogenous residential sorting may also attenuate the OLS estimates if households with stronger preferences for segregated neighborhoods disproportionately select neighborhoods more likely to adopt covenants, even if those neighborhoods ultimately did not adopt these covenants.

Robustness: The baseline results in Table B.1 use radian sectors of $r = 10$, $z_j = 1000$ meters, and the number of houses built by 1930 as a measure of the 1940 urbanization

[6] Non-White households include Black, Asian, and Native American households, as defined by the census, regardless of their country of birth.

Figure B.4: Distribution of Covenant Shares across Developments

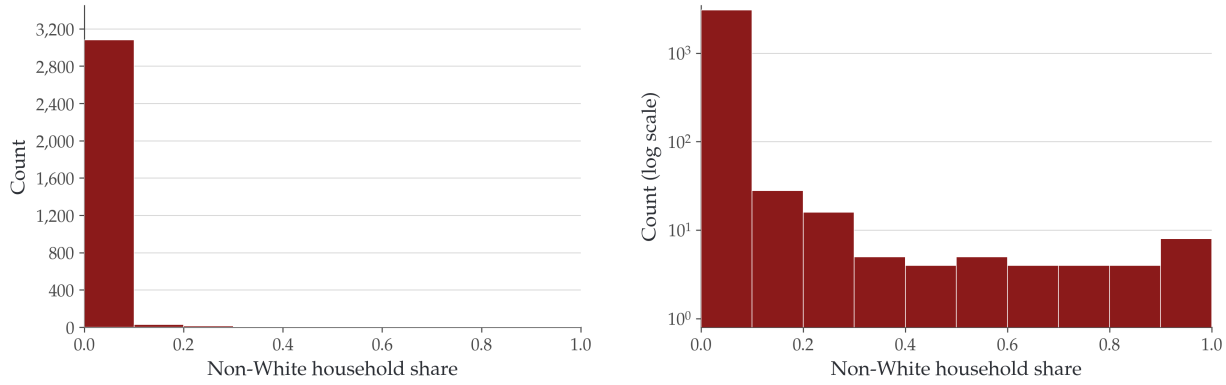


Note: This figure illustrates the distribution of covenants shares across developments, presented as counts on the left and on a logarithmic scale on the right.

index. These findings remain robust when considering a longer time period (houses built by 1910; Appendix Table B.2, Panel A) and a varying buffer zone that decreases as the distance from the CBD increases ($z_j = 500 - 2000$ meters; see Appendix Table B.2, Panel B). There is also a concern that our findings might be influenced by the Home Owners' Loan Corporation (HOLC) redlining (Aaronson et al., 2021). However, we believe this is unlikely because the covenants predate HOLC, and HOLC only covered parts of Minneapolis, not the entire metro area. To ensure that HOLC redlining does not affect our results, we control for the share of D-graded areas in developments, and the baseline results remain unchanged (Appendix Table B.2, Panel C). Furthermore, there may be a concern that our results are driven by the 3.89% of households we could not match to specific developments. When we make alternative assumptions about matching these households to different developments, the estimated results remain consistent with the baseline findings (see Appendix Table B.2, Panel D). Finally, in the baseline analysis, we do not control for household characteristics because they could be endogenously correlated with racial covenants. However, controlling for the head of household's occupation, ownership status (owner, renter, or living in group quarters), marital status, or the number of total and working members in the household, does not alter the baseline effects of covenants on racial and ethnic sorting (see Appendix Table B.2, Panel E).[7]

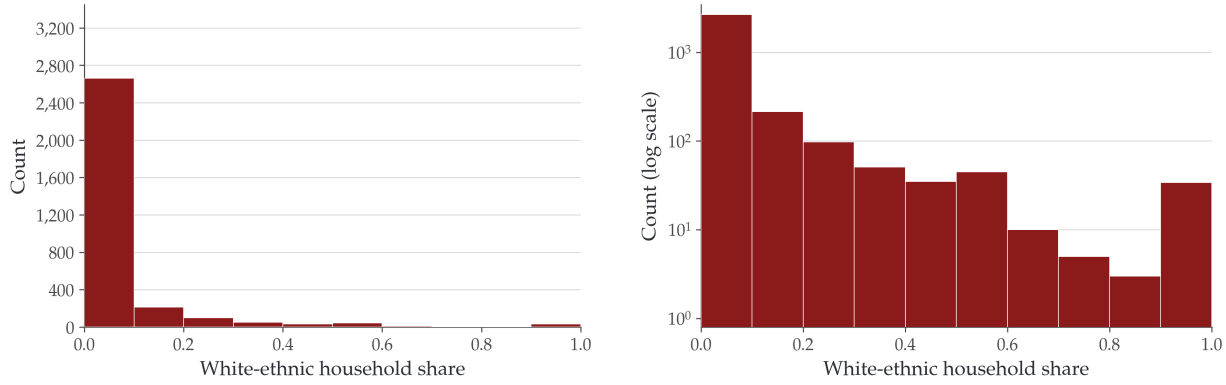
[7]We do not control for housing prices or rents in our reduced-form specifications because these prices are themselves influenced by racial covenants and neighborhood composition making them an endogenous

Figure B.5: Distribution of Minority Shares across Developments



(a) Non-White

(b) Non-White, logarithmic scale



(c) White-ethnic

(d) White-ethnic, logarithmic scale

Note: This figure illustrates the distribution of non-White and White-ethnic household shares across developments, presented as counts on the left and on a logarithmic count scale on the right. White-ethnic includes White Eastern and Southern European immigrants and White immigrants in the Other category. Shares are calculated over all households within each development. Each of the 3,158 developments is counted once, including developments with zero group share.

Figure B.4 plots the distribution of covenants shares across developments, presented as counts on the left and on a logarithmic scale on the right. Figure B.5 plots the distribution of non-White and White-ethnic shares across developments, presented as counts on the left and on a logarithmic scale on the right.

control; controlling for them could introduce bias. Additionally, since census only collected income data from a 5% sample, we use the occupation indicator as a proxy for household income and wealth. For further details, see Appendix **??**.

Table B.2: Racial Covenants and Sorting: Robustness

	Non-White	White NE Imm.	White WE Imm.	White EE/SE Imm.	White Other Imm.
Panel A: 1910 - 1940 Timeframe					
Share Covenanted	-1.149***	0.065	0.419**	-0.092	-0.556
	(0.353)	(0.768)	(0.180)	(0.109)	(0.507)
Observations	164,673	164,673	164,673	164,673	164,673
First-Stage F-Stat	22.73	22.73	22.73	22.73	22.73
Panel B: Instrument with Shrinking Buffer Zone					
Share Covenanted	-0.133***	-0.087	0.066**	-0.026*	-0.046
	(0.041)	(0.165)	(0.027)	(0.015)	(0.093)
Observations	164,673	164,673	164,673	164,673	164,673
First-Stage F-Stat	94.44	94.44	94.44	94.44	94.44
Panel C: Baseline + Additional Control for Share with D-Grade Rating in HOLC					
Share Covenanted	-0.101***	-0.050	0.047**	-0.028**	-0.005
	(0.034)	(0.148)	(0.023)	(0.013)	(0.083)
Observations	164,673	164,673	164,673	164,673	164,673
First-Stage F-Stat	90.08	90.08	90.08	90.08	90.08
Panel D: Alternative Household to Development Matching					
Share Covenanted	-0.099***	-0.030	0.048**	-0.026**	-0.003
	(0.034)	(0.146)	(0.023)	(0.013)	(0.082)
Observations	164,673	164,673	164,673	164,673	164,673
First-Stage F-Stat	92.59	92.59	92.59	92.59	92.59
Panel E: Baseline + Additional Household-Level Controls					
Share Covenanted	-0.102***	-0.082	0.046**	-0.031**	-0.044
	(0.036)	(0.144)	(0.024)	(0.013)	(0.083)
Observations	164,670	164,670	164,670	164,670	164,670
First-Stage F-Stat	90.64	90.64	90.64	90.64	90.64
Mean Dep. Variable	0.010	0.151	0.005	0.003	0.064
Location controls	Yes	Yes	Yes	Yes	Yes
ED FEs	Yes	Yes	Yes	Yes	Yes

Note: This table plots the instrumental variable (IV 2SLS) parameter estimates from Equation B.1. Non-White includes households identified as Black, Asian, or Native American, regardless of birth country. White Northern European (NE), Western European (WE), Eastern and Southern European (EE/SE), and Other Immigrants (Imm.) refer to White households from these regions (see Appendix B for details). Robust standard errors in parentheses. All panels control for location variables such as log slope, and log distances to the central business district (CBD), heavy-industry zoning, commercial zoning, the nearest lake, and the streetcar. Additionally, they control for development-level mean log lot size and mean build year. ED FEs are enumeration-district (ED) fixed effects. Panel A employs a longer pre-period (1910-1940) for the urban frontier instrument. Panel B constructs the instrument with a shrinking buffer zone based on distance to the CBD (from 2000 to 500 meters). Panel C adds a control for the share of development with D-grade rating in HOLC redlining. Panel D uses an alternative household-to-development matching for 3.89% of households. Panel E adds household controls of ownership tenure, occupation indicator, marital status, and number of household members and working members to the baseline specification. * < 0.1, ** < 0.05, *** < 0.01.

References

Aaronson, Daniel, Daniel Hartley, and Bhashkar Mazumder, "The Effects of the 1930s HOLC "Redlining" Maps," *American Economic Journal: Economic Policy*, 2021, *13* (4), 355–392.

Brigham, Carl C., *A Study of American Intelligence*, Princeton, NJ: Princeton University Press, 1923.

Danbom, David B., "Flour Power: The Significance of Flour Milling at the Falls," *Minnesota History*, 2003, *58* (5/6), 270–285.

Hoyt, Homer, *One Hundred Years of Land Values in Chicago: The Relationship of the Growth of Chicago to the Rise in Its Land Values, 1830–1933*, Chicago: University of Chicago Press, 1933.

Irwin, Douglas A., *Peddling Protectionism: Smoot–Hawley and the Great Depression*, Princeton, NJ: Princeton University Press, 2011.

Pyle, Lizbeth A., "The Land Market beyond the Urban Fringe," *Geographical Review*, 1985, *75* (1), 32–43.

Sood, Aradhya and Kevin Ehrman-Solberg, "The Long Shadow of Housing Discrimination: Evidence from Racial Covenants," *The Review of Economic Studies*, 2026.

C. Model Details

C.1 Choice Sets under Common and Match-Specific Admission Draws

This appendix shows that the nested choice sets of Section 3.1.2 are a special case of a general admission rule. It then extends the model to admission draws that vary across neighborhoods and shows how to compute match probabilities in that case.

C.1.1 A General Admission Rule

For a household of group k, define the admission index of neighborhood j as $a_j^k \equiv \alpha^k X_j^s$ and its demand index as $v_j(k) \equiv \exp\big(\delta_j + \Delta\beta^k X_j^d\big)$. Suppose that neighborhood j admits household i if and only if

$$a_j^{k(i)} \;\geq\; \eta_{ij},$$

where the vector of draws $\boldsymbol{\eta}_i = (\eta_{i1}, \ldots, \eta_{iJ})$ has joint distribution $F_{\boldsymbol{\eta}}$. The model of Section 3.1.2 is the special case in which the draw is common across neighborhoods, $\eta_{ij} = \eta_i$ for all j. Each realization $\boldsymbol{\eta}$ generates a choice set

$$\mathcal{O}(\boldsymbol{\eta}; k) \equiv \{j \in \mathcal{J} :\; a_j^k \geq \eta_j\}.$$

For any choice set $\mathcal{O} \subseteq \mathcal{J}$, let

$$L_j(\mathcal{O}; k) \equiv \frac{v_j(k)\,\mathbb{1}\{j \in \mathcal{O}\}}{1 + \sum_{\ell \in \mathcal{O}} v_\ell(k)}, \qquad L_0(\mathcal{O}; k) \equiv \frac{1}{1 + \sum_{\ell \in \mathcal{O}} v_\ell(k)}$$

denote the logit probabilities of choosing neighborhood j and the outside option when the choice set is fixed at $\mathcal{O}$. These are the conditional choice probabilities in (2). The match probability integrates over the admission draws:

$$\mathbb{P}(y_{ij} = 1 \mid k) = \int L_j\big(\mathcal{O}(\boldsymbol{\eta}; k); k\big)\, dF_{\boldsymbol{\eta}}(\boldsymbol{\eta}) = \sum_{\mathcal{O} \subseteq \mathcal{J}} q_{\mathcal{O}}(k)\, L_j(\mathcal{O}; k), \tag{C.1}$$

where $q_{\mathcal{O}}(k) \equiv \mathbb{P}\big(\mathcal{O}(\boldsymbol{\eta}_i; k) = \mathcal{O}\big)$ is the probability that the choice set is exactly $\mathcal{O}$. The second equality holds because the sets $\{\boldsymbol{\eta} : \mathcal{O}(\boldsymbol{\eta}; k) = \mathcal{O}\}$ partition the support of $\boldsymbol{\eta}$, and the integrand equals $L_j(\mathcal{O}; k)$ on each of them. The outside-option probability replaces L_j

with L_0. Expression (C.1) is the match probability (5): households draw a latent choice set, then choose by logit within it.

C.1.2 Common Draw: Nested Choice Sets

Proposition 1 (Common admission draw)**.** *Suppose* $\eta_{ij} = \eta_i \sim \mathcal{N}(0,1)$ *for all* j*. Order neighborhoods so that* $a^k_{(1)} \geq \cdots \geq a^k_{(J)}$*, and let* $\mathcal{O}^n \equiv \{(1), \ldots, (n)\}$ *for* $n = 0, \ldots, J$*. Then* $q_{\mathcal{O}}(k) > 0$ *only for the* $J+1$ *nested sets* $\mathcal{O}^n$*, with*

$$q_{\mathcal{O}^n}(k) = \Phi\big(a^k_{(n)}\big) - \Phi\big(a^k_{(n+1)}\big), \qquad \Phi\big(a^k_{(0)}\big) \equiv 1, \quad \Phi\big(a^k_{(J+1)}\big) \equiv 0,$$

and the match probability is

$$\mathbb{P}(y_{ij} = 1 \mid k) = \sum_{n=0}^{J} \Big[\Phi\big(a^k_{(n)}\big) - \Phi\big(a^k_{(n+1)}\big)\Big] L_j\big(\mathcal{O}^n; k\big). \tag{C.2}$$

Proof. A household with draw η is admitted to neighborhood (n) if and only if $\eta \leq a^k_{(n)}$. Because the indices are ordered, a household admitted to (n) is admitted to every neighborhood with a higher index. A draw in the interval $\big(a^k_{(n+1)},\, a^k_{(n)}\big]$ therefore generates exactly the choice set $\mathcal{O}^n$, and this interval has probability $\Phi(a^k_{(n)}) - \Phi(a^k_{(n+1)})$. These $J+1$ intervals partition the real line, so every other subset has probability zero. Substituting into (C.1) gives (C.2). □

Let $\text{rank}_k(j)$ denote the position of neighborhood j in this ordering. Because $L_j(\mathcal{O}^n; k) = 0$ for $n < \text{rank}_k(j)$, the sum in (C.2) runs over the sets that contain j. Sorting the neighborhoods once per group and accumulating $\sum_{\ell \leq n} v_{(\ell)}(k)$ evaluates all match probabilities in $O(J \log J)$ operations. The empty set, $n = 0$, has probability $1 - \Phi(a^k_{(1)})$. Households who draw it are admitted nowhere and choose the outside option.

C.1.3 Extension: Match-Specific Draws

A common draw restricts choice sets to be nested. A household admitted to a hard-to-enter neighborhood is admitted to every neighborhood that is easier to enter. Match-specific draws relax this restriction. They allow, for example, a developer-specific component of screening that a common household draw cannot capture.

Proposition 2 (Match-specific admission draws)**.** *Suppose $\boldsymbol{\eta}_i$ has joint distribution $F_{\boldsymbol{\eta}}$. Then every subset $O \subseteq \mathcal{J}$ can have positive probability, with*

$$q_O(k) = \int_{R_O(k)} dF_{\boldsymbol{\eta}}(\boldsymbol{\eta}), \qquad R_O(k) \equiv \left\{\boldsymbol{\eta} : \ \eta_j \le a_j^k \text{ for } j \in O, \ \ \eta_j > a_j^k \text{ for } j \notin O\right\},$$

and the match probability is (C.1)*, a mixture over all 2^J choice sets. If the draws are independent, $\eta_{ij} \overset{iid}{\sim} \mathcal{N}(0,1)$, then*

$$q_O(k) = \prod_{j \in O} \Phi\big(a_j^k\big) \prod_{j \notin O} \Big[1 - \Phi\big(a_j^k\big)\Big].$$

Proof. The regions $R_O(k)$ are exactly the sets of draws that generate choice set O, and they partition the support of $\boldsymbol{\eta}$. The first expression follows from (C.1). Under independence, each neighborhood admits the household with probability $\Phi(a_j^k)$, independently of the others, which gives the product form. □

With J in the thousands, summing over 2^J choice sets is infeasible. Under independent draws, the match probability reduces to a one-dimensional integral. Write $p_j \equiv \Phi(a_j^k)$ and use the identity $1/D = \int_0^\infty e^{-sD}\, ds$ for $D > 0$. Taking expectations over the independent admission indicators,

$$\mathbb{P}(y_{ij} = 1 \mid k) = \int_0^\infty e^{-s}\, v_j(k)\, p_j\, e^{-s\, v_j(k)} \prod_{\ell \neq j} \Big[1 - p_\ell + p_\ell\, e^{-s\, v_\ell(k)}\Big] ds,$$

$$\mathbb{P}(y_{i0} = 1 \mid k) = \int_0^\infty e^{-s} \prod_{\ell \in \mathcal{J}} \Big[1 - p_\ell + p_\ell\, e^{-s\, v_\ell(k)}\Big] ds.$$

Each integral is one-dimensional and smooth, so standard quadrature evaluates it with $O(J)$ operations per node. The likelihood and the contraction for δ_j then carry over unchanged, with these probabilities in place of (C.2). Correlated draws, for instance a household component plus a match-specific component, nest both cases. They can be handled by simulating the household component and applying the integral above conditional on it.

C.2 Ignoring Access Restrictions: Proofs and Covariance Representations

This appendix proves Theorem 1 and Corollary 1, and states two related representations. As in the toy model of Section 3, utility is $U_{ij} = \beta_0 X_j + \epsilon_{ij}$, choice sets O_i are nonempty and drawn independently of the taste shocks, the outside option is suppressed, and X_j is not constant on $\mathcal{J}$. For any β, define the full-set logit weights and the corresponding moments

$$w_j(\beta) \equiv \frac{e^{\beta X_j}}{\sum_{\ell \in \mathcal{J}} e^{\beta X_\ell}}, \qquad \mathbb{E}_{w(\beta)}[g_j] \equiv \sum_{j \in \mathcal{J}} w_j(\beta)\, g_j, \qquad \mathrm{Var}_{w(\beta)}(X_j) \equiv \mathbb{E}_{w(\beta)}[X_j^2] - \big(\mathbb{E}_{w(\beta)}[X_j]\big)^2,$$

so that $m_{\mathcal{J}}(\beta) = \mathbb{E}_{w(\beta)}[X_j]$. Let $\pi_j \equiv \mathbb{E}_i\big[e^{\beta_0 X_j}\mathbb{1}\{j \in O_i\}/\sum_{\ell \in O_i} e^{\beta_0 X_\ell}\big]$ be the true probability that a household chooses j. Then $\sum_j \pi_j X_j = \mathbb{E}_i\big[m_{O_i}(\beta_0)\big]$.

Proof of Theorem 1. The population objective of the misspecified logit is

$$Q(\beta) = \sum_{j \in \mathcal{J}} \pi_j \log w_j(\beta).$$

Differentiating,

$$Q'(\beta) = \sum_{j \in \mathcal{J}} \pi_j X_j - m_{\mathcal{J}}(\beta) = \mathbb{E}_i\big[m_{O_i}(\beta_0)\big] - m_{\mathcal{J}}(\beta), \qquad Q''(\beta) = -\frac{d\, m_{\mathcal{J}}(\beta)}{d\beta} = -\,\mathrm{Var}_{w(\beta)}(X_j) < 0.$$

The objective is strictly concave, so the pseudo-true coefficient $\widetilde{\beta}$, whenever it is finite, is the unique solution of $m_{\mathcal{J}}(\widetilde{\beta}) = \mathbb{E}_i[m_{O_i}(\beta_0)]$. A finite solution exists whenever $\mathbb{E}_i[m_{O_i}(\beta_0)]$ lies strictly between $\min_j X_j$ and $\max_j X_j$, because $m_{\mathcal{J}}(\beta)$ is continuous and approaches these bounds as $\beta \to -\infty$ and $\beta \to \infty$. Subtracting $m_{\mathcal{J}}(\beta_0)$ from both sides,

$$m_{\mathcal{J}}(\widetilde{\beta}) - m_{\mathcal{J}}(\beta_0) = \Gamma.$$

By the mean value theorem, the left-hand side equals $\mathrm{Var}_{w(\beta^*)}(X_j)\,(\widetilde{\beta} - \beta_0)$ for some β^* between $\widetilde{\beta}$ and β_0. Dividing by the variance, which is strictly positive, gives the expression

for $\widetilde{\beta} - \beta_0$ and the sign result. □

Proof of Corollary 1. Say that restrictions remove high-X_j neighborhoods if every choice set has the form $O_i = \{j \in \mathcal{J} : X_j \leq c_i\}$, with $O_i \neq \mathcal{J}$ with positive probability. For any such set $O \neq \mathcal{J}$, let $\lambda_O \equiv \sum_{j \in O} w_j(\beta_0) \in (0,1)$. Then

$$m_{\mathcal{J}}(\beta_0) = \lambda_O \, m_O(\beta_0) + (1 - \lambda_O) \, m_{\mathcal{J} \setminus O}(\beta_0).$$

Every neighborhood outside O has a value of X_j above every neighborhood inside it, so $m_{\mathcal{J} \setminus O}(\beta_0) > m_O(\beta_0)$ and hence $m_O(\beta_0) < m_{\mathcal{J}}(\beta_0)$. For $O = \mathcal{J}$, the two are equal. Taking expectations over choice sets gives $\Gamma < 0$, and Theorem 1 implies $\widetilde{\beta} < \beta_0$. If restrictions instead remove low-X_j neighborhoods, $O_i = \{j : X_j \geq c_i\}$, the same argument gives $\Gamma > 0$ and $\widetilde{\beta} > \beta_0$. □

The corollary requires only that the excluded neighborhoods have higher (lower) values of X_j than the accessible ones. It does not require that households agree on the cutoff. The down-payment example in Section 3 satisfies this condition, with household-specific cutoffs c_i.

Remark 1 (Common choice set: covariance representation)**.** *Suppose every household faces the same choice set, $O_i = A \subsetneq \mathcal{J}$, and let $r_j \equiv \mathbb{1}\{j \in A\}$. Then $m_A(\beta_0) = \mathbb{E}_{w(\beta_0)}[X_j r_j] / \mathbb{E}_{w(\beta_0)}[r_j]$, so the selection gap and the bias are*

$$\Gamma = \frac{\mathrm{Cov}_{w(\beta_0)}(X_j, r_j)}{\mathbb{E}_{w(\beta_0)}[r_j]}, \qquad \widetilde{\beta} - \beta_0 = \frac{\mathrm{Cov}_{w(\beta_0)}(X_j, r_j)}{\mathbb{E}_{w(\beta_0)}[r_j] \; \mathrm{Var}_{w(\beta^*)}(X_j)}.$$

The sign of the bias is the sign of the logit-weighted covariance between the characteristic and access. The weights emphasize the neighborhoods that matter most in the unrestricted market.

Proposition 3 (Multivariate representation)**.** *Suppose utility is $U_{ij} = X_j' \beta_0 + \epsilon_{ij}$ with $X_j \in \mathbb{R}^K$, and define $m_{\mathcal{S}}(\beta)$ as the vector of logit-weighted means of X_j over $\mathcal{S}$. Let $\Gamma \equiv \mathbb{E}_i[m_{O_i}(\beta_0)] - m_{\mathcal{J}}(\beta_0)$ and*

$$\Sigma(\beta) \equiv \sum_{j \in \mathcal{J}} w_j(\beta) \big(X_j - m_{\mathcal{J}}(\beta)\big) \big(X_j - m_{\mathcal{J}}(\beta)\big)'.$$

If $\bar{\Sigma} \equiv \int_0^1 \Sigma\big(\beta_0 + t(\widetilde{\beta} - \beta_0)\big)\, dt$ *is nonsingular, then*

$$\widetilde{\beta} - \beta_0 = \bar{\Sigma}^{-1}\,\Gamma.$$

With a common choice set A, $\Gamma = \mathrm{Cov}_{w(\beta_0)}(X_j, r_j)/\mathbb{E}_{w(\beta_0)}[r_j]$.

Proof. The gradient of the population objective is $\nabla Q(\beta) = \mathbb{E}_i[m_{O_i}(\beta_0)] - m_{\mathcal{J}}(\beta)$, and its Hessian is $-\Sigma(\beta)$. At the pseudo-true coefficient, $\nabla Q(\widetilde{\beta}) = 0$. The integral form of the mean value theorem gives $0 = \nabla Q(\beta_0) - \bar{\Sigma}\,(\widetilde{\beta} - \beta_0)$, and $\nabla Q(\beta_0) = \Gamma$. Rearranging yields the result. □

With several characteristics, the sign of a given component of $\widetilde{\beta} - \beta_0$ need not equal the sign of the corresponding component of Γ, because $\bar{\Sigma}^{-1}$ mixes regressors.

Remark 2 (Selection on unobservables)**.** *The results above concern observed characteristics. A direct analogue of selection on unobservables in Dubé et al. (2021) arises if the admission rule is enriched to* $\alpha^{k(i)} X_j^s + \omega_j \geq \eta_i$*, where* ω_j *is an unobserved access shifter. Then* $\mathrm{Cov}(\omega_j, \xi_j) \neq 0$ *means that access is selected on unobserved demand.*

C.3 Monte Carlo Illustration

Figure C.1 illustrates Theorem 1 and Corollary 1. There are $J = 200$ neighborhoods with $X_j \sim \mathcal{N}(0,1)$, drawn once and held fixed, and $N = 50{,}000$ households with $\beta_0 = 1$. A share ρ of neighborhoods is removed from every household's choice set. In the left panel, the removed neighborhoods are those with the highest X_j. In the right panel, they are those with the lowest X_j. Households choose by logit within the accessible set, and we vary ρ from 0 to 0.8 with 100 replications each. For each ρ, we estimate a correctly specified logit on the accessible set and a misspecified logit on the full set. The solid line is the pseudo-true coefficient from Theorem 1, computed without simulation.

The correctly specified estimator recovers β_0 at every ρ. The misspecified estimator tracks the pseudo-true coefficient. When restrictions remove high-X_j neighborhoods, the estimated taste falls below β_0 and becomes negative once about a quarter of neighborhoods is removed. When restrictions remove low-X_j neighborhoods, the estimated taste rises above β_0.

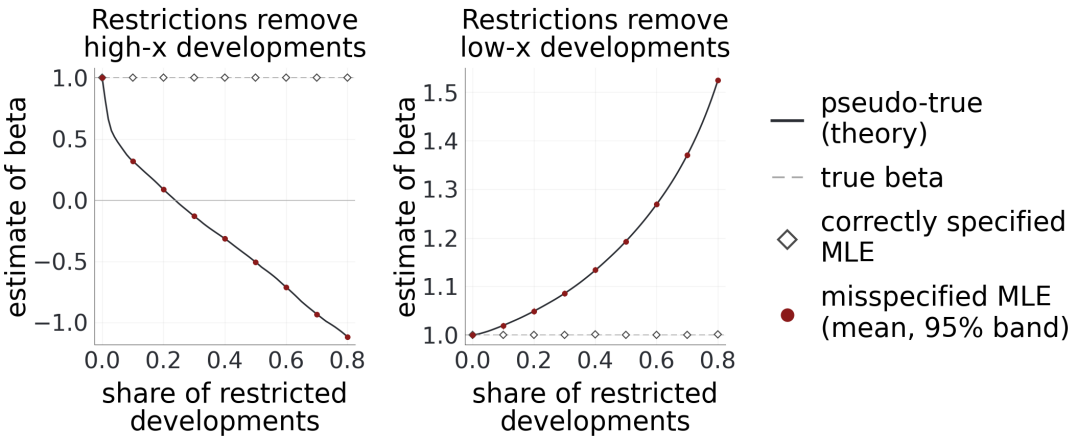


Figure C.1: Monte Carlo: Estimated Taste When Choice Sets Are Ignored

Note: The horizontal axis is the share of neighborhoods removed from every household's choice set. Dots and bands show the mean and 95% band of the misspecified estimates across 100 replications. Diamonds show the correctly specified estimates. The solid line is the pseudo-true coefficient and the dashed line is the true coefficient, $\beta_0 = 1$.

References

Dubé, Jean-Pierre, Ali Hortaçsu, and Joonhwi Joo, "Random-Coefficients Logit Demand Estimation with Zero-Valued Market Shares," *Marketing Science*, 2021, *40* (4), 637–660.

D. Identification and Estimation Details

D.1 Excluded Shifter: Developer Leniency

We construct developer leniency in three steps. First, for each development, we compute the minority household share across the developer's *other* developments. This leave-one-out share excludes the development itself. Second, we require at least five households in those other developments. Third, we code leniency as one if this share is above its median across locations with an observed share, and zero otherwise.

Leniency is missing in three cases: the developer is unknown, the developer built a single development, or the other developments hold fewer than five households. In these cases, we set leniency to zero and include a separate missing indicator in the admission rule.

Table D.1 reports balance tests. We regress each characteristic on log price, the covenant share, and log development size, and report household-weighted mean residuals. The left panel compares developments whose developer is above the median with those at or below it. The right panel compares developers with a single development with those with several. Of the 24 differences, none is significant at the 5% level. Three are significant at the 10% level: the 1910 immigrant share and the distance to the nearest lake in the left panel, and the 1910 Black share in the right panel.

D.2 Identification Intuition

This appendix shows how the excluded shifter separates admission from preferences. The argument uses derivatives of the log match probability.

Removing the mean utility. Consider a minority group k. Non-minority households are admitted everywhere, so their match probability is the standard logit,

$$\log \mathbb{P}(y_{ij} = 1 \mid nm) = \delta_j - \log\left(1 + \sum_{\ell \in \mathcal{J}} e^{\delta_\ell}\right).$$

Table D.1: Developer Balance Tests

	Propensity to admit			Single- versus multiple-addition developers		
Characteristic	At or below median	Above median	t-stat.	Single addition	Multiple additions	t-stat.
1910 ED Black household share residual (pp)	−0.15	0.13	0.87	−0.18	0.07	1.95*
1910 ED immigrant household share residual (pp)	−2.1	1.8	1.66*	−0.5	0.2	0.42
1910 ED occupational score residual	0.12	−0.10	−0.53	−0.19	0.08	0.94
1910 ED homeownership share residual (pp)	0.1	−0.1	−0.07	0.1	0.0	−0.04
Lot size residual (sq ft)	−2,692	2,296	1.31	−2,699	1,096	1.48
East–west location residual (km; positive = east)	−0.11	0.09	0.58	−0.18	0.07	1.11
North–south location residual (km; positive = north)	−0.27	0.23	1.35	0.17	−0.07	−0.78
Plat year residual (years)	0	0	0.39	0	0	−0.41
Build year residual (years)	0	0	0.37	1	0	−1.18
Share platted after 1910 residual (pp)	0.2	−0.2	−0.11	0.9	−0.4	−0.50
Distance to lake residual (km)	−0.12	0.10	1.69*	−0.04	0.02	0.65
Distance to streetcar residual (km)	0.00	0.00	0.09	−0.01	0.01	0.64
Addition–EDs	811	865	–	969	1,676	–
Distinct additions	607	607	–	793	1,214	–

Notes: Entries are household-weighted mean residuals after controlling for log price, covenant share, and log development size. In the left panel, the t-statistic tests above-median minus at-or-below-median developers. In the right panel, it tests multiple-development minus single-development developers. Tests use household-count analytic weights and standard errors clustered by developer. Population-share residuals are in percentage points. Year residuals are deviations in years. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

For group k, Equation (6) gives

$$\log \mathbb{P}\big(y_{ij} = 1 \mid k\big) = \delta_j + \Delta\beta^k X_j^d + \log G_j^k, \qquad G_j^k \equiv \sum_{n \geq \text{rank}_k(j)} \frac{\Phi\big(a_{(n)}^k\big) - \Phi\big(a_{(n+1)}^k\big)}{1 + \sum_{\ell \leq n} v_{(\ell)}(k)}.$$

Define the log ratio of the two match probabilities, $r_j^k \equiv \log \mathbb{P}(y_{ij} = 1 \mid k) - \log \mathbb{P}(y_{ij} = 1 \mid nm)$. Then

$$r_j^k = \Delta\beta^k X_j^d + \log G_j^k + c,$$

where c does not vary across neighborhoods. The mean utility δ_j cancels, and so does the demand shock ξ_j. Any characteristic that all groups value equally drops out, including a common valuation of leniency or covenants.

A single admission index. The term G_j^k depends on neighborhood j in two ways. It depends on the admission index a_j^k through the choice-set probabilities. It also depends on the demand index $v_j(k)$ through the denominators. Hold the other neighborhoods fixed, and consider changes in j's covariates that leave the ordering unchanged. Summation by parts gives

$$G_j^k = \frac{\Phi\big(a_j^k\big)}{1 + S_r^k} + \sum_{n > r} \Phi\big(a_{(n)}^k\big)\Big[\frac{1}{1 + S_n^k} - \frac{1}{1 + S_{n-1}^k}\Big], \qquad S_n^k \equiv \sum_{\ell \leq n} v_{(\ell)}(k),$$

with $r = \text{rank}_k(j)$. Suppose that $v_j(k)$ is small relative to $1 + S_n^k$ for every set that contains j. This holds when each neighborhood is a small part of every choice set that contains it. The denominators then do not depend on j's covariates. G_j^k then depends on these covariates only through a_j^k. We can write

$$r_j^k \approx \underbrace{\Delta\beta^k X_j^d}_{\text{preferences}} + \underbrace{\varphi_j^k\big(a_j^k\big)}_{\text{choice sets}} + c, \qquad \varphi_j^k(a) \equiv \log\Big(\frac{\Phi(a)}{1 + S_r^k} + C_j^k\Big),$$

where C_j^k collects the second sum above.

Derivatives. To fix ideas, let w_j be the excluded shifter and p_j the log price. The shifter enters the admission index but not $\Delta\beta^k X_j^d$. The price enters both. The relevant terms are

$a_j^k = \alpha_w w_j + \alpha_p p_j + \cdots$ and $\Delta\beta^k X_j^d = \Delta\beta_p p_j + \cdots$, where we drop the superscript k on the coefficients. Differentiating r_j^k gives

$$\frac{\partial r_j^k}{\partial w_j} = \varphi' \alpha_w, \qquad \frac{\partial^2 r_j^k}{\partial w_j^2} = \varphi'' \alpha_w^2,$$
$$\frac{\partial r_j^k}{\partial p_j} = \Delta\beta_p + \varphi' \alpha_p, \qquad \frac{\partial^2 r_j^k}{\partial p_j \, \partial w_j} = \varphi'' \alpha_p \alpha_w.$$

The ratio of the two second derivatives gives the relative weight of price in the admission index:

$$\frac{\alpha_p}{\alpha_w} = \frac{\partial^2 r_j^k / \partial p_j \, \partial w_j}{\partial^2 r_j^k / \partial w_j^2}.$$

Subtracting the admission response from the price response gives the preference for price:

$$\Delta\beta_p = \frac{\partial r_j^k}{\partial p_j} - \frac{\alpha_p}{\alpha_w} \frac{\partial r_j^k}{\partial w_j}.$$

The second step then recovers β_p^{nm}, so $\beta_p^k = \beta_p^{nm} + \Delta\beta_p$.

What each assumption does. The argument uses five assumptions.

1. The common draw η_i makes admission a function of a single index, $\varphi_j^k(a_j^k)$.
2. The index has no interaction between p_j and w_j. The ratio α_p/α_w is then a constant that does not depend on φ_j^k.
3. The function φ_j^k is nonlinear, $\varphi'' \neq 0$. Otherwise, both second derivatives are zero and the ratio is not defined.
4. The shifter w_j does not enter $\Delta\beta^k X_j^d$. Otherwise, $\partial r_j^k/\partial w_j = \Delta\beta_w + \varphi' \alpha_w$, and the last step recovers $\Delta\beta_p - (\alpha_p/\alpha_w)\Delta\beta_w$ instead of $\Delta\beta_p$.
5. The shifter varies enough to compute both derivatives in w_j.

The argument does not use the shape of φ_j^k. It also does not require utility to be linear in p_j.

Limitations. The argument is local. It holds the rest of the market fixed and uses a small-neighborhood approximation. In the data, we compare different neighborhoods,

and φ_j^k differs across them through S_r^k and C_j^k. The maximum likelihood estimator uses the full structure of (6) rather than these derivatives. In addition, developer leniency is binary in our baseline specification. The derivatives in w_j then become a comparison of neighborhoods with and without a lenient developer at the same price. This discrete comparison relies more on the functional form of φ_j^k.

Match-specific draws. The same argument applies with independent draws η_{ij} (Appendix C.1). When $v_j(k)$ is small, the integral in Proposition 2 gives $\mathbb{P}(y_{ij} = 1 \mid k) \approx v_j(k)\,\Phi(a_j^k)\,H^k$, where H^k does not depend on j's covariates. Then $\varphi_j^k(a) = \log \Phi(a)$, which is the same for every neighborhood and strictly concave.

Without the excluded shifter. Without an excluded shifter, the logit and normal functional forms still identify the parameters. A model with full admission and logit errors satisfies the independence of irrelevant alternatives. Departures from it in the data then reveal the formation of choice sets, as in Abaluck and Adams-Prassl (2021). Identification then rests entirely on these functional forms. The excluded shifter reduces this reliance.

D.3 Instruments

Table D.2 presents the first-stage regressions for the second step of the estimation. Each endogenous variable is regressed on both instruments.

D.3.1 Shift-Share Instrument Construction

To address the potential endogeneity in minority household location decisions, we construct a shift-share instrument similar to Card (2001) and Tabellini (2020). The instrument combines three key components: growth rates from non-Minnesota metropolitan areas, historical migrant settlement patterns across enumeration districts (EDs), and immigration quotas from the US Immigration Act of 1924. We use the 1930 and 1940 IPUMS censuses to construct the instrument. We first apportion 1930 household counts to 1940 ED boundaries using land-area overlap weights. We then calculate the predicted 1940 number of households for each ED, e, using the average growth rate of non-Minnesota metropolitan areas between 1930 and 1940:

$$\hat{H}_{e,1940} = H_{e,1930}(1 + g_{1930-1940}) \tag{D.1}$$

Table D.2: First Stage for the Second Step of Estimation

	(A) Log rent	(B) Non-minority share
Terrain slope	0.070** (0.033)	0.119*** (0.036)
Shift–share	0.172*** (0.038)	−0.124*** (0.030)
Cragg–Donald F		43.71
Kleibergen–Paap F		5.94
Lewis–Mertens		19.40
critical value ($\tau = 0.10$, $\alpha = 0.05$)		45.38
Coarse ED fixed effects	✓	✓
Location controls	✓	✓
Locations		3,937

Notes: This table shows the first-stage regressions for the instruments used in the second step of the two-sided matching model's estimation. Observations are locations, weighted by their number of households. Standard errors in parentheses. All specifications include coarse ED fixed effects and controls for log distances to the central business district, streetcar, heavy industry, commercial areas, and the nearest lake, mean log lot size per household, and log development size. The Lewis–Mertens critical value corresponds to a 10% worst-case bias threshold at the 5% level. Significance levels: *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

where $H_{e,1930}$ is the apportioned 1930 household count and $g_{1930–1940}$ represents the average growth rate across non-Minnesota metropolitan areas.

For each sending international region $r \in R$, immigrant group $k \in K$, and ED e, we calculate historical settlement patterns using 1930 data:

$$\omega_{erk} = \frac{M_{erk,1930}}{\sum_{j \in E} \sum_{\ell \in K} M_{jr\ell,1930}} \tag{D.2}$$

where $M_{erk,1930}$ represents the apportioned number of households with a white immigrant head from region r in group k in ED e in 1930, and E denotes the set of 1940 EDs covered by the crosswalk. The denominator pools immigrant groups within each sending region.

We then predict the number of new immigrant households in each ED by combining these settlement patterns with immigration quotas:

$$\hat{M}_{ek,1940} = \sum_{r} \omega_{erk} \cdot Q_r \cdot 9 \tag{D.3}$$

where Q_r represents the annual quota for region r under the US Immigration Act of 1924, multiplied by 9 to cover each year of the 1930–1938 period. We carry out this calculation separately for white minority immigrants and other white immigrants. We then construct the ED-level instrument as:

$$Z_{ek,1940} = \frac{M_{ek,1930} + \hat{M}_{ek,1940}}{\hat{H}_{e,1940}} \tag{D.4}$$

where $M_{ek,1930}$ represents the existing immigrant household stock in ED e in 1930 for group k. The instrument for all white immigrants is the sum of the instruments for the two groups.

For development-level analysis, we merge this ED-level instrument to development–ED observations within ED e:

$$Z_{aek,1940} = Z_{ek,1940}. \tag{D.5}$$

The instrument is constant across developments within an ED. This instrument satisfies the exclusion restriction insofar as national immigration quotas were set without reference

to local neighborhood or metro area economic and social conditions, and pre-period settlement patterns from 1930 are uncorrelated with subsequent unobserved development-level changes once development and location characteristics are controlled for.

References

Abaluck, Jason and Abi Adams-Prassl, "What Do Consumers Consider before They Choose? Identification from Asymmetric Demand Responses," *The Quarterly Journal of Economics*, 2021, *136* (3), 1611–1663.

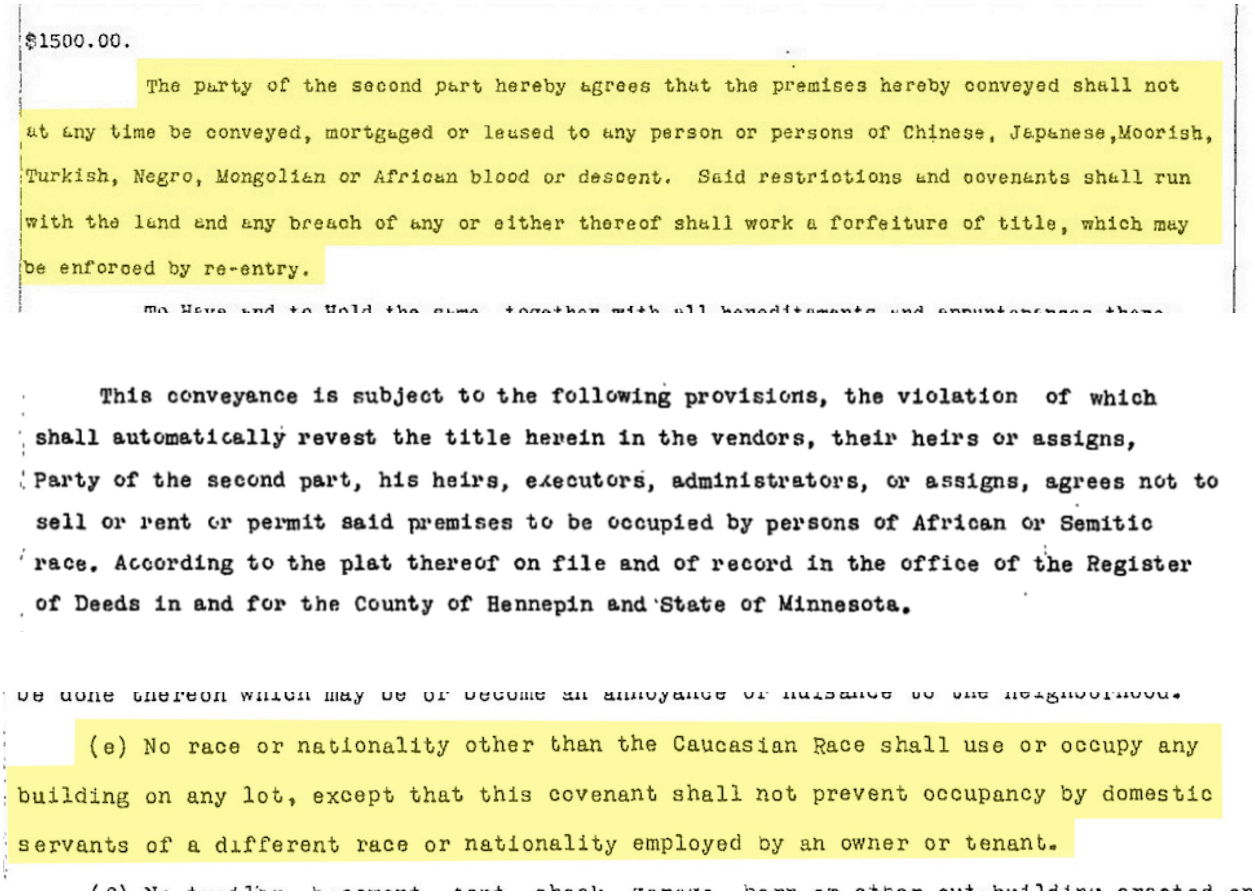
Card, David, "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration," *Journal of Labor Economics*, 2001, *19* (1), 22–64.

Tabellini, Marco, "Gifts of the Immigrants, Woes of the Natives: Lessons from the Age of Mass Migration," *The Review of Economic Studies*, 2020, *87* (1), 454–486.

E. Additional Results

F. Additional Figures

Figure F.1: Examples of Racial Covenants



$1500.00.

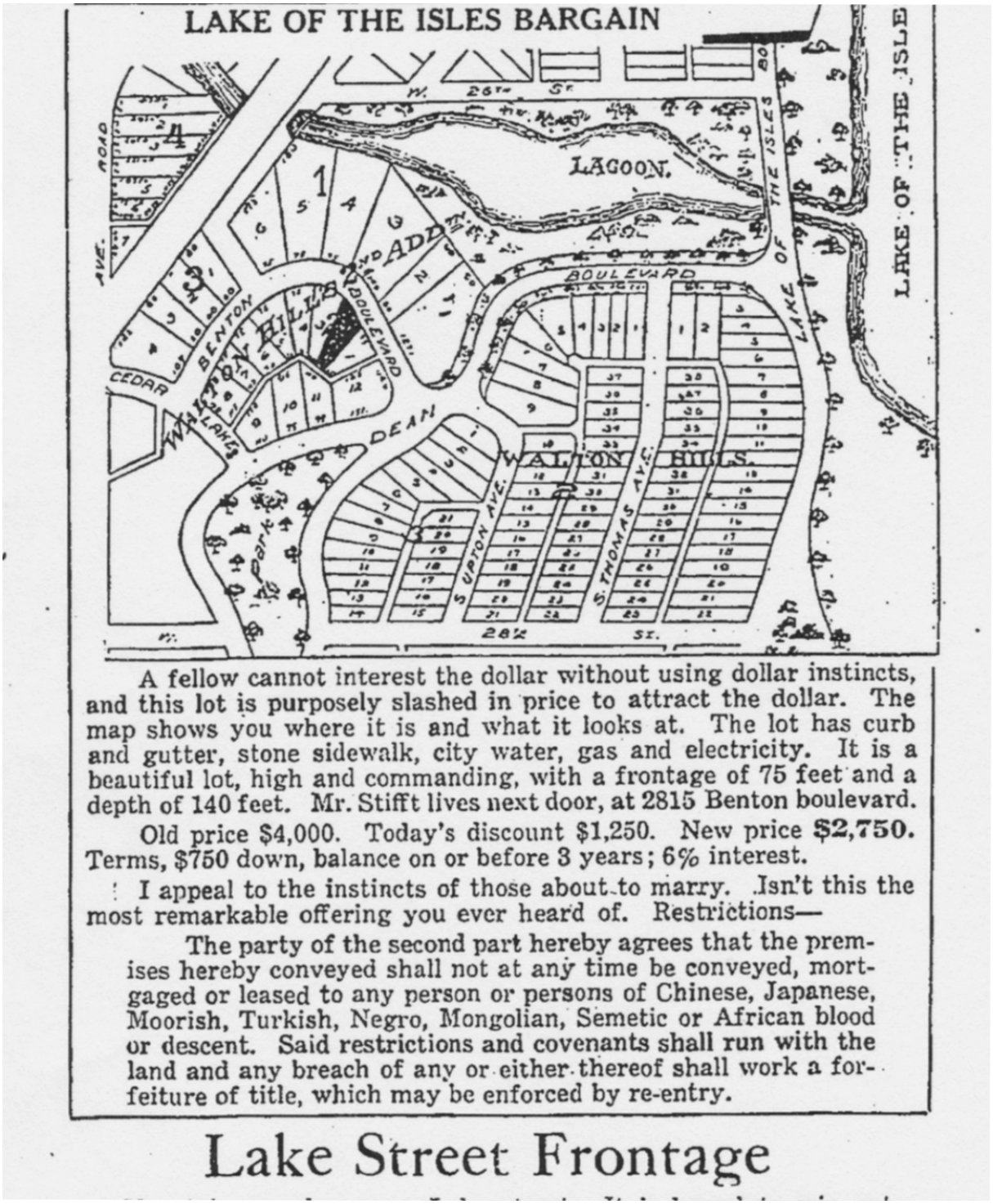
The party of the second part hereby agrees that the premises hereby conveyed shall not at any time be conveyed, mortgaged or leased to any person or persons of Chinese, Japanese, Moorish, Turkish, Negro, Mongolian or African blood or descent. Said restrictions and covenants shall run with the land and any breach of any or either thereof shall work a forfeiture of title, which may be enforced by re-entry.

This conveyance is subject to the following provisions, the violation of which shall automatically revest the title herein in the vendors, their heirs or assigns, Party of the second part, his heirs, executors, administrators, or assigns, agrees not to sell or rent or permit said premises to be occupied by persons of African or Semitic race. According to the plat thereof on file and of record in the office of the Register of Deeds in and for the County of Hennepin and State of Minnesota.

(e) No race or nationality other than the Caucasian Race shall use or occupy any building on any lot, except that this covenant shall not prevent occupancy by domestic servants of a different race or nationality employed by an owner or tenant.

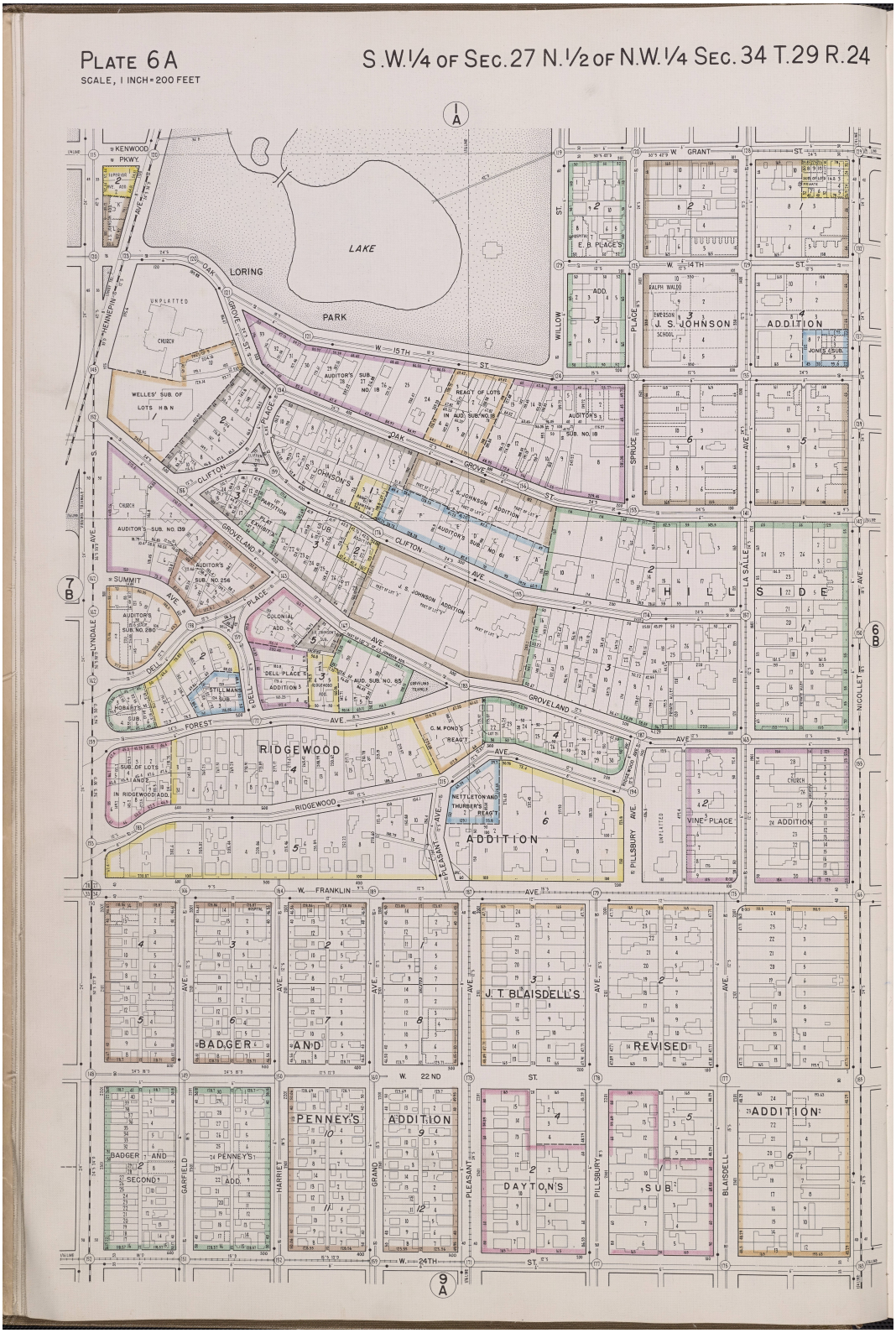
Note: This figure provides three examples of racial covenants in sales deeds in Hennepin County, Minnesota.

Figure F.2: Newspaper Advertisement of Racial Covenants in a New Development



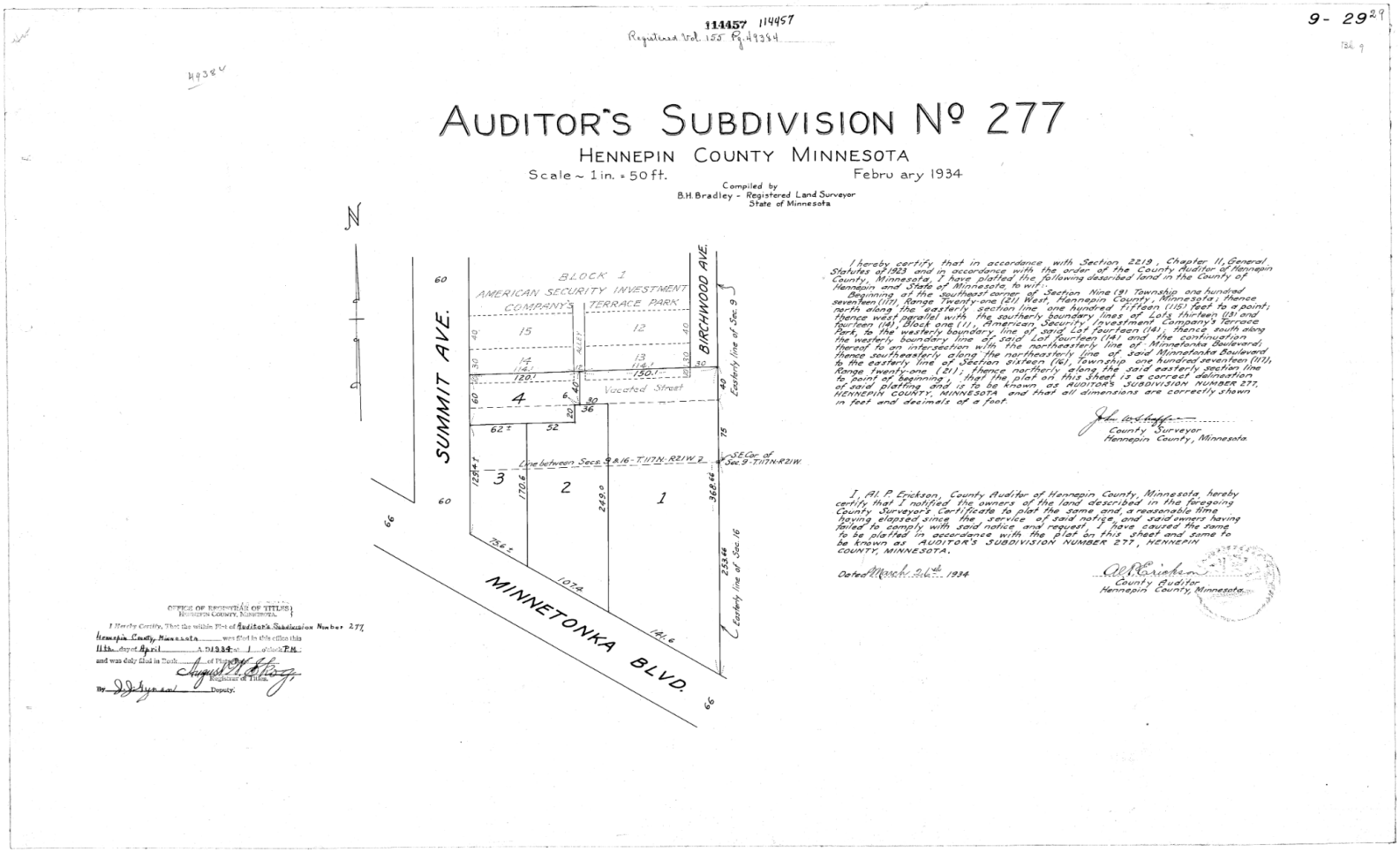
Note: This figure provides an example of newspaper advertisement of racial covenants in the 'Walton Hills' development, *Minneapolis Morning Tribune, January 12, 1919*.

Figure F.3: Example Page from the 1940 Minneapolis City Atlas



Note: This figure provides an example page from the 1940 Minneapolis City Atlas, showing to-scale real estate development boundaries and development names.

Figure F.4: Example Development Plat Map



Note: This figure reproduces the plat map for Auditor's Subdivision No. 277 in the Minneapolis metro area. Development plat maps record subdivision boundaries and lots, the subdivision year, and the developer or development firm. We digitize these records to construct subdivision years and link developers across developments.