Who Marries Whom?
The Role of Segregation by Race and Class*

Benjamin Goldman, Harvard University †
Jamie Gracie, Harvard University
Sonya R. Porter, U.S. Census Bureau

March 2024

Click here for most recent version

Abstract

Americans rarely marry outside of their race or class group. We distinguish between two possible explanations: a lack of exposure to other groups versus a preference to marry within group. We develop an instrument for neighborhood exposure to opposite-sex members of other race and class groups using variation in sex ratios among nearby birth cohorts in childhood neighborhoods. We then test whether increased exposure results in more interracial (white-Black) and interclass (top-to-bottom parent income quartile) marriages. Increased exposure to opposite-sex members of other class groups generates a substantial increase in interclass marriage, but increased exposure to other race groups has no detectable effect on interracial marriage. We use these results to estimate a spatial model of the marriage market and quantify the impact of reducing residential segregation in general equilibrium. For small changes in exposure, the model implies effects in line with recent estimates from policy experiments. We then use the model to assess the overall contribution of segregation and find that residential segregation has large effects on interclass, but not interracial, marriage.

*Any opinions and conclusions expressed herein are those of the authors and do not represent the views of the U.S. Census Bureau. The Census Bureau has ensured appropriate access and use of confidential data and has reviewed these results for disclosure avoidance protection (Project 7519874: CBDRB-FY23-CES014-009, CBDRB-FY23-0492). This research was funded by Opportunity Insights, the Stone Program at Harvard Kennedy School, the National Science Foundation (Graduate Research Fellowship, grant DGE-2140743), and Harvard University.

†Job Market Paper. Email: bgoldman@g.harvard.edu. I thank my advisors, Raj Chetty, Claudia Goldin, Nathan Hendren, Larry Katz, and Jesse Shapiro, for their support, mentorship, and patience. Feedback from Kaveh Danesh, Brad Foster, João Loula, Andrei Shleifer, and Matt Staiger was essential to this project. We also want to thank Isaiah Andrews, Zach Bleemer, Brian Bushard, Sahil Chinoy, Will Dobbie, John Friedman, Ed Glaeser, Jeff Gortmaker, Abigail Hiller, Martin Koenen, Lauren Russell, Dani Sandler, Ben Scuderi, Elie Tamer, Winnie van Dijk, Rob Watson, and Joseph Winkelmann.

I Introduction

Americans rarely marry outside of their race or class group (Bloome and Ang, 2020; Charles, Hurst and Killewald, 2013). In 2019, only 13% of newly married white Americans married a non-white spouse, despite 40% of the U.S. population being non-white. Similarly, only 8% of newly married Americans from high-income (top parental income quartile) families married a spouse from a low-income (bottom quartile) family. The fragmentation of families along race and class lines is central to many issues of longstanding societal concern, such as the dynamics of income across generations, discrimination, and the transfer of norms from parents to children (Althoff, Gray and Reichardt, 2022; Bisin and Verdier, 2001; Bobo, 2001; Chadwick and Solon, 2002; Mare, 2000; Olivetti et al., 2020).

Why do people have such a strong tendency to marry within their own race and class group? We distinguish between two possible explanations for the tendency to marry within group, known as marital homophily: lack of *exposure*, i.e., social interactions with other groups are less likely, versus *preferences*, meaning a desire to marry someone from your own group, even conditional on exposure.[1] Distinguishing between these explanations has important implications for policy. For instance, efforts to reduce residential segregation will have much larger effects on marital patterns—and, consequently, on the dynamics of income across generations—if exposure is primarily responsible for marital homophily.

In this paper, we estimate the effect of exposure on marital homophily, focusing specifically on the impacts of reducing residential segregation. We do this in three steps. First, we present descriptive facts on the extent of marital homophily and segregation by race and class. Since the impact of reduced segregation on homophily depends on whether proximity impacts one's choice of marital partners, we also present descriptive statistics on the role of geography in marital sorting. The second part of the paper estimates, in partial equilibrium, whether individuals are willing to marry across race and class lines. To do this, we leverage sex ratio variation in childhood neighborhoods, which generates variation in exposure to opposite-sex members of other race and class groups. In the final section, we build and estimate a spatial model of the marriage market to predict the effect of reducing segregation in general equilibrium, where changes to the marriage market in one neighborhood can spill over across space. Our findings suggest that marital homophily by class is driven, in large part, by residential segregation, whereas racial homophily is not.

We conduct this analysis using administrative data that links the universe of federal tax records

[1] We use the term "preferences" to describe the propensity for within group marriage conditional on exposure. These "preferences" need not be specific to the individual or innate in anyway. They can reflect preferences of family or community, and they can arise endogenously from life experiences and social norms that can, in principle, be affected by exposure (Allport, Clark and Pettigrew, 1954; Barlow et al., 2012; Corno, La Ferrara and Burns, 2022; Paolini, Harwood and Rubin, 2010). We return to a discussion of this in Section V.

to the 2000 and 2010 decennial censuses. We capture race and ethnicity in the decennial censuses and measure class using parental household income during childhood. We use "high-income" and "low-income" as shorthands for parents with incomes in the top and bottom quartiles of the household income distribution, and "white" and "Black" as shorthands for non-Hispanic white and non-Hispanic Black. Our baseline estimates focus on individuals in the 1982-1989 birth cohorts and measure marital status at age 30, when 35% of people are married. We include additional estimates at later ages for a subset of cohorts.

We focus our analysis on race and class for two primary reasons. First, our measures of race and class are not impacted by marriage outcomes or marriage market conditions.[2] Second, marriage plays a central role in the race- and class-based economic disparities that are salient and persistent features of American society (Althoff, Gray and Reichardt, 2022; Bayer and Charles, 2018; Chadwick and Solon, 2002; Chetty et al., 2014; Eriksson et al., 2023; Mare, 2000). Prior estimates show that marriage rates and spousal incomes account for as much as 85% of the adult household income disparity between white and Black children who grew up in households with the same family income (Binder et al., 2022; Chetty et al., 2020). Motivated by this fact, our main results on race focus on white-Black marriage.

Rates of interclass and interracial marriage are low (Charles, Hurst and Killewald, 2013; Ermisch, Francesconi and Siedler, 2006; Fryer, 2007; Olivetti et al., 2020). In our data, we find 3.1% of people from a high-income family have a spouse from a low-income family at age 30, and only 0.5% of white individuals have a Black spouse. Our sample includes unmarried people to avoid conditioning on the selected segment of the population that chooses to marry. We quantify the level of homophily by comparing these rates to two benchmarks. In the first benchmark, everyone marries at the population marriage rate and chooses a spouse at random. In the second, group-specific marriage rates are held fixed and spouses are randomly reshuffled among the existing pool of married people. The observed rates of interclass and interracial marriage are one-third to one-half as large as these benchmarks, implying high levels of marital homophily by class and race.

The impact of segregation on homophily depends on the extent to which neighbors are an important component of the marriage market. Neighborhoods are relevant because they provide meeting places and can determine the people with whom one works and socializes (Athey et al., 2021; Chetty et al., 2022; McPherson, Smith-Lovin and Cook, 2001). In a recent survey, 35% of

[2]Individuals may change there racial identification, but this is rare, especially for white and Black individuals in our sample period (Liebler et al., 2017). Our measure of class is invariant because it is based on parental, rather than own, income. Many papers studying marital homophily focus on the role that homophily on education and individual income has on inequality in household income (e.g., Eika, Mogstad and Zafar, 2019; Greenwood et al., 2014; Pencavel, 1998; Mare, 1991). However, more recent work has documented the endogeneity of education and income to the marriage market which can complicate their use as measures of homophily (Calvo, Lindenlaub and Reynoso, 2021; Chiappori, Dias and Meghir, 2018; Chiappori, Iyigun and Weiss, 2009; Cullen and Gruber, 2000; Lafortune, 2013; Low, 2021; Lundberg, 1985).

online daters said they limit their search to within their neighborhood (Kirkham, 2019). Consistent with this, 68% of married couples lived within 50 census tracts of each other five years prior to marriage.[3] In contrast, the average share of the U.S. population that lives within 50 census tracts of a given person is 0.07%.[4] Marriage probabilities also decay sharply with distance. The probability of marrying someone who grew up in the same childhood census tract is 10 times greater than the probability of marrying someone in the 50th closest tract.

Having established that people tend to marry nearby neighbors, we then turn to documenting the composition of their neighborhoods. Much prior work has shown that the U.S. is segregated by race and income (e.g., Hall, Iceland and Yi, 2019; Logan et al., 2020; Massey and Denton, 1988; Reardon and Bischoff, 2011). We construct a measure of exposure to other groups based on neighbors in the nearest 50 census tracts. We upweight individuals who are geographically closer using weights chosen based on how marriage probabilities vary with distance. Using this measure, we show that neighborhoods where individuals live as young adults, which we define as ages 18 to 27, are highly segregated by race and class. We take this as evidence that the segregation documented in prior literature remains high for a definition of neighborhood that comprises a large fraction of the marriage market.

These facts show that people tend to marry nearby neighbors, and that their neighbors tend to look like them. But what if their neighbors look different? To study the role of exposure in explaining marital homophily, we begin by documenting the association between rates of interracial and interclass marriage and exposure to other groups in young adulthood. For individuals from low-income families, a 10 percentage point (pp) increase in the share of their neighbors who come from high-income families is associated with a 2.2pp increase in the probability of being married to a spouse from a high-income family. Among Black individuals, a 10pp increase in the share of their neighbors who are white is associated with a 0.9pp increase in the probability of having a white spouse. We conclude that there is substantially less marital homophily in integrated neighborhoods than in segregated neighborhoods, though the association between segregation and cross-group marriage is stronger for class than race.

The association between homophily and neighborhood integration could be driven by selection. For example, people from low-income families who live in high-income neighborhoods as adults might have unobserved traits that lead them to prefer or be preferred by a spouse from a high-income family. If this were the case, the association between homophily and segregation

[3] This falls to 58% when excluding couples already living with each other. If we look 10 years prior to marriage and exclude couples already living with each other, we find 45% of couples have lived within 50 census tracts of each other a decade before they are married. Even considering childhood neighborhoods, which we define as the parental neighborhood of a child at age 18, 45% of marriages are between two people who grew up within 50 census tracts of each other.

[4] Census tracts are a small geographic unit constructed by the U.S. Census Bureau. Tracts are designed to have approximately $4,000$ residents; in 2010, there were $74,134$ tracts in the U.S.

would not represent the causal effect of exposure.

To properly disentangle the causal effect of exposure from selection, we construct an instrument for exposure based on race- and class-specific sex ratios in childhood neighborhoods.[5] The instrument exploits differences in market "tightness," which we define as the difference in the fraction of opposite- versus own-sex neighbors from a given race or class group. For example, going from a market where 10% of male and female neighbors are Black to one where 5% of male neighbors are Black and 15% of female neighbors are Black would represent a 10pp increase in Black market tightness for men. Zaremba (2022) uses sex ratio variation at birth to estimate the impact of bargaining power in the marriage market on maternal health outcomes for Black women. We refrain from estimating effects on downstream outcomes and concentrate on the direct impacts of shifts in market tightness on marriage outcomes.

Our strategy relies on three empirical facts: 1) 99% of legal marriages in the U.S. are opposite-sex marriages (Kreider et al., 2023); 2) it is common to marry a neighbor who is close in age; and 3) people tend to live as young adults near where they grew up (Foster, 2017; Sprung-Keyser, Hendren and Porter, 2022).[6] The first fact motivates our definition of tightness as opposite sex share relative to own sex share. The second fact defines cells of nearby neighborhoods and birth cohorts that are important for marriage prospects but small enough to generate sex ratio variation. The third generates a first stage by inducing a correlation between the demographic composition of child and adult neighborhoods.

The intuition for the instrument can be seen with a simple hypothetical. Suppose a male child is born in a place and time where, by chance, the nearby high-income families had more female than male children. As an adult, he will tend to be in a marriage market where there are relatively more females than males from high-income families. Although the variation in sex ratios across neighborhoods and cohorts is small, the scale of our data allows us to use this variation to estimate the partial-equilibrium effect of market tightness on interclass and interracial marriage.

Our instrumental variable (IV) estimates show that increased tightness—that is, the difference in the fraction of opposite-sex versus own-sex neighbors between ages 18 and 27 who come from other race and class groups—causes a substantial increase in interclass marriage. For people from low-income families, a 10pp increase in market tightness among individuals from high-income

[5]The first wave of studies using sex ratios to study marriage markets used observational variation in adult neighborhoods (Cox, 1940; Easterlin, 1961; Groves and Ogburn, 1928; South and Lloyd, 1992). More recent studies tend to use policy changes (Goni, 2022), or quasi-experiments that impact sex ratios (Abramitzky, Delavande and Vasconcelos, 2011; Angrist, 2002; Bazzi et al., 2023; Brainerd, 2007; Charles and Luoh, 2010; Hoxby, 2000). We leverage the longitudinal nature of our data to hone in on sex ratios in childhood neighborhoods, which we show come from random sex assignment at birth rather than sorting to neighborhoods.

[6]The Supreme Court ruling that legalized same-sex marriage nationwide, Obergefell v. Hodges, occurred in 2015, the middle of our sample period. As a result, marriage, as measured in tax data during our sample period, is almost exclusively opposite-sex marriage.

families leads to a 1.3pp increase in the probability of an interclass marriage. For people from high-income families, a 10pp increase in market tightness among individuals from low-income families leads to an increase of 1.4pp. These estimates show that increased market tightness across class lines generates interclass marriages.

The IV estimates for race show a strikingly different result. For white individuals, a 10pp increase in market tightness among Black individuals results in only a 0.0002pp increase in the probability of interracial marriage. For Black individuals, a 10pp increase in market tightness among white individuals results in a 0.0623pp increase in the probability of interracial marriage. These estimates have standard errors of 0.197 and 0.374, allowing us to clearly reject exposure effects of the same magnitude as those we found for interclass marriage. Increased market tightness across race lines does not seem to increase interracial marriage.

We use multiple strategies to evaluate the validity of our IV estimates. We find that the distribution of the instrument matches an idealized counterpart—constructed by treating birth sex as an IID Bernoulli random variable—extremely closely. This would not be true if parents sorted into neighborhoods based on the sex of their child. Furthermore, the instrument is unrelated to observable characteristics of neighborhoods and families. Within a neighborhood and group, the sex ratios across years are unrelated, and within a neighborhood and year, the sex ratios of different race and class groups are unrelated. We also show that our IV estimates are robust to including controls such as family fixed effects, where we compare siblings who are exposed to different levels of market tightness due to chance variation in sex ratios around the time of their birth.

We next turn to evaluating how changes in residential segregation would affect marriage patterns. Our IV estimates alone are not sufficient to estimate marital patterns under counterfactual levels of segregation. There are two main reasons for this. First, reducing segregation involves holding market tightness—the share of a given group who are male versus female—fixed, while changing market thickness, the overall share of a given group in a neighborhood. The sex ratio instrument generates variation in market tightness, but not in market thickness. Second, our instrument estimates the impact of changes in market tightness on marriage outcomes in a given neighborhood (partial equilibrium). But neighborhoods are not closed-marriage markets. Anything that changes marital outcomes in one neighborhood necessarily affects outcomes in others, because of the spillover effects through the matching market. Estimating the total change in marriage outcomes requires keeping track of spillovers across neighborhoods (general equilibrium).

To address these limitations, we build a spatial model of the marriage market. We start from the perfectly transferable utility matching model of Choo and Siow (2006), but relax two commonly used assumptions. The first is the assumption that the U.S. can be treated as a single large marriage market (e.g., Chiappori, Oreffice and Quintana-Domeque, 2012, 2018; Mourifié and Siow, 2021; Siow, 2015). In our context, this assumption is rejected by the fact that individuals disproportionately marry their nearby neighbors. We relax this assumption by including neighborhood in the type space and allowing individuals to have preferences over distance from their potential spouse.[7]

The second assumption is that there are no unobserved traits that correlate with the observed traits, meaning any observed relationship between intergroup marriage and segregation is a causal effect. Both a priori intuitions—that residents of neighborhoods with different compositions are likely to be different in unobserved ways—and the evidence, such as the distinction between the OLS and IV estimates of the effect of exposure on interracial marriage, suggest that this assumption is not a good approximation in our setting. We model traits that are unobserved to the researcher but can correlate with neighborhood choice and use our IV results to estimate model parameters net of selection effects. This method can be applied in other matching applications where omitted traits are a potential concern and a credible instrument is available.[8]

We estimate the model using simulated method of moments (McFadden, 1989). The moments we include in our estimation are the national marriage outcomes by race and class, the association between marriage outcomes and neighborhood exposure to other race and class groups, and the IV effects of market tightness. We simulate these moments in the model and find the set of model parameters that minimize the distance between the simulated moments and the data moments.

We use our model to estimate the impact of segregation on marital homophily under two types of counterfactuals. In both counterfactuals, we estimate the impact of segregation while holding preferences fixed—in the future, our framework can be extended to allow preferences to be endogenous to exposure. In the first counterfactual, we simulate the impact of reducing segregation by moving a small share of the overall population. This type of change has been studied in the context of Moving to Opportunity, Hope VI, and the Gautreaux Project (Bergman et al., 2019; Chyn, Collinson and Sandler, 2023; Ludwig et al., 2013; Popkin, 2004). In our model, the impact of these moves on intergroup marriage depends on the distance cost and preferences for marrying within (as opposed to across) race and class. As distance costs increase, marriage markets become spatially smaller, and the impact of reducing segregation on intergroup marriage increases. As preferences to marry within (versus across) race and class increase, the effect of reducing segregation decreases. At our estimated model parameters, these policies have large effects (per person who moves) on interclass and interracial marriage.

---

[7] Adding census tract expands the type space of the model substantially, adding substantial computational complexity. We show how to re-express the equilibrium as a system of quadratic equations which can be easily solved with fixed point methods. We plan to make our code publicly available to others working on high-dimensional matching problems.

[8] Our approach is related to recent work that moves beyond the Gumbel distribution to allow for richer forms of unobserved preference heterogeneity (Galichon and Salanié, 2022; Graham, 2013; Gualdani and Sinha, 2020; Fox, 2010; Sinha, 2018). Rather than positing a richer distribution, we allow for a second unobservable component of preferences and use an instrument unrelated to the unobservable type to support identification and estimation.

In the case of class, for people from bottom- and top-quartile families, the model implies that a 10pp increase in the share of neighbors who are from the other class group increases rates of interclass marriage by 1.7pp. This relationship is close to the gradient between interclass marriage and cross-class exposure across U.S. commuting zones, where a 10pp increase in exposure is associated with a 1.8pp increase in interclass marriage. The model-simulated impact of reducing segregation matches the actual difference in rates of interclass marriage between segregated and integrated cities.

In the case of race, for white and Black individuals, the model implies that a 10pp increase in the share of neighbors from the other race group increases rates of interracial marriage by 1.2pp. The Gautreaux Project provides a natural comparison for these estimates. The program gave Black families in high-poverty neighborhoods, chosen at random, a housing voucher to move to low-poverty and predominantly white suburbs. Chyn, Collinson and Sandler (2023) estimate the impact of these moves on rates of interracial marriage and finds that a 10pp increase in the share of neighbors who are white increases rates of interracial marriage by 1.7pp.

Having found that the model provides a good match to the experimental effects of small changes, we turn to using the model to study the effect of a larger change. In our second counterfactual, we remove the role of distance altogether, thereby eliminating any impact of segregation on the marriage market. The model suggests that eliminating segregation would do little to increase rates of interracial marriage. Without any distance costs, the share of Black people with a white spouse would increase only slightly from 2.1% to 2.5%.

Comparing the effect of the first counterfactual—moving a small number of people—to the effect of the second counterfactual—eliminating the role of distance entirely—is instructive about the forces in the model and, we believe, in the real-world marriage market. In a large population with heterogeneous preferences, some individuals will prefer an interracial marriage even when average preferences for interracial marriage are low. High levels of segregation can prevent these interracial matches from forming, but small reductions in segregation generate enough of an increase in cross-race exposure for these matches to form. Subsequent larger reductions in segregation generate few incremental interracial marriages. This is consistent with the across commuting zone data, where even the most integrated commuting zones have only slightly higher rates of interracial marriage than highly segregated commuting zones.

Our model can help to reconcile what, at first glance, might seem to be contradictory findings in prior work. In addition to the Gautreaux Project, cross-race exposure has been shown to increase rates of interracial relationships in other contexts (e.g., Corno, La Ferrara and Burns, 2022; Gordon and Reber, 2018; Merlino, Steinhardt and Wren-Lewis, 2019). At the same time, other studies have documented strong preferences against interracial marriage (Fisman et al., 2008; Hitsch, Hortaçsu and Ariely, 2010; Wong, 2002). We show that these patterns are consistent. In a population with heterogeneous preferences, low *average* preferences for interracial marriage, and high levels of segregation, small increases in cross-race exposure can generate meaningful increases in interracial marriage.

Unlike interracial marriage, removing the effects of distance generates a large increase in interclass marriage. At baseline, 3.1% of children from low-income families go on to have a spouse from a high-income family, compared with 19.0% for children who themselves are from high-income families. This gap of 15.9pp falls to 9.3pp (a reduction of 41.5%) in a hypothetical world without distance costs. These estimates are consistent with a large fraction of the population being open to an interclass marriage. As a result, segregation prevents interclass marriages from forming by restricting exposure across class lines.

Marital homophily is an important component of inequality within (Ciscato and Weber, 2020; Clark and Cummins, 2022; Eika, Mogstad and Zafar, 2019; Schwartz, 2010) and across generations (Althoff, Gray and Reichardt, 2022; Buckles et al., 2023; Keller and Shiue, 2023; Ermisch, Francesconi and Siedler, 2006; Espín-Sánchez, Ferrie and Vickers, 2023). Despite similarly low rates of interclass and interracial marriage, the factors underlying race- and class-based marital homophily are not the same. Residential segregation is responsible for nearly half of homophily by class, but only a small fraction of homophily by race. Our results show that the marriage market is an important channel through which segregation can impact economic inequality.

The paper is organized as follows: Section II describes the data; Section III presents descriptive evidence on marital homophily, neighborhood segregation and the role of distance in partner choice; and Section IV estimates the causal effect of exposure to opposite-sex members of other groups on interclass and interracial marriage. In Section V, we build a spatial model of the marriage market and predict the effect of reducing segregation in general equilibrium. Section VI concludes.

II Data

We share our analysis sample with Chetty et al. (2023), which extends the data in Chetty et al. (2020). We rely on three primary sources: the 2000 and 2010 decennial censuses, federal tax returns from 1979, 1984, 1989, 1994, 1995, 1998-2019, and the American Community Survey (ACS) from 2005-2019. The datasets are linked using a unique person identifier called a Protected Identification Key (PIK). PIKs are assigned using information including Social Security Numbers (SSNs), names, addresses and dates of birth. The record linkage process is described in detail in Wagner and Lane (2014). We limit our analysis to individuals who are assigned a PIK. The data are stripped of personally identifiable information.

II.A Sample Definitions

We begin our sample construction by finding all individuals who were ever claimed as a dependent on a tax return. We match the individual and their claimer(s) to the 2020 Numident file to assign a year of birth. The claimer is a potential parent if they were between the ages 15 and 50 in the year the child was born. The first potential parent(s) who claim the individual are assigned as the invariant parent(s), regardless of any subsequent changes in marital status or dependent claiming.[9] This procedure relies on matching to the Numident file, a dataset covering all people with SSNs, and will therefore exclude any individuals who are unauthorized immigrants to the U.S. or who were claimed by unauthorized immigrants.

We measure parental household income during childhood in our sample of linked children and parents. Following Chetty et al. (2023), we construct mean parent income between ages 13 and 17 based on the parent(s)' tax returns in those years. We use Adjusted Gross Income in the years where parents file a tax return and impute household income using W-2 income for non-filers. The W-2 data begin in 2005 and we code non-filers as having zero income if the W-2 data are unavailable.[10] We restrict to families with positive income since having negative income is typically due to capital losses, which is a sign that the family is wealthy (Chetty et al., 2014). The remaining individuals form our parent and child linked data.

We limit our parent- and child-linked data to individuals born between 1978 and 1999. Claiming rates begin to drop at age 16 (Chetty et al., 2014)—the 1978 birth cohort turns 16 in 1994, which is our first year of dependent claiming data. Our primary analysis sample consists of individuals born between 1982 and 1989. We use parental income data to measure the class of both individuals and their spouse. By beginning with the 1982 cohort, we can measure spouse class for the 86% of married individuals who have a spouse who is less than 5 years older than them. Our baseline estimates use marriage outcomes at age 30, which requires us to end our sample with the 1989 birth cohort since our last year of tax data is 2019.

II.B Variable Definitions

In this subsection, we briefly define the variables we use in our primary analysis. We measure all monetary variables in 2015 dollars, adjusting for inflation using the consumer price index (CPI-U).

Class. Our measure of class is based on parent income during childhood. We measure parental income in each year using the total pre-tax income at the household level, which we label household

[9] A person who claims a dependent on a tax return does not have to be that individual's biological parent, but they do have to be financially supporting the individual. We restrict those ages to 15-50 to avoid links to grandparents or siblings. Chetty et al. (2020) show that for children claimed in the year 2000, 93% live with those same parents in the 2000 decennial census.

[10] Chetty et al. (2020) show that the median income reported in the ACS for non-filers is $5,000.

income. We use the term household income for simplicity, but we do not include incomes from cohabitating partners or other household members aside from the primary tax filer's spouse. In years where a parent files a tax return, we define household income as the sum of Adjusted Gross Income, social security payments and tax-exempt interest payments, as reported on their 1040 tax return. For non-filers, we define household income as W-2 income when available. Otherwise, we consider household income for non-filers to be zero. We define parent income as mean household income over the five years in which the child is age 13 through 17. We construct percentile ranks of parent income by ranking parents relative to other parents with children in the same birth cohort. For much of our analysis, we split the sample into quartiles and refer to parents in the bottom and top quartile of the parental income distribution as "low-" and "high-income" parents.

Spouse Class. For individuals who are married, spouse class is defined analogously to own class using the same measure of parent income during childhood. Spouse class is missing if the spouse cannot be linked to parents. This group includes spouses born prior to 1978 and those who did not spend their childhood in the U.S., as well as people who fail to meet any of the criteria used in the construction of the linked parent and child data described above. We are able to assign spouse class for 85% of married individuals in our primary analysis sample. When measuring marriage outcomes, we assign those with missing spouse class to an "out-of-sample" class category.

Race. We assign each individual to a race and ethnicity group using information from the U.S. Census Bureau and the ACS. We take the responses in priority order from the 2010 short form, the 2000 short form, and the 2005-2019 ACS surveys. Following Chetty et al. (2020), we focus on the following race and ethnicity groups: non-Hispanic white, non-Hispanic Black, non-Hispanic Asian, non-Hispanic American Indian and Alaskan Native (AIAN), Hispanic and non-Hispanic "other." The last group contains Native Hawaiian and Other Pacific Islanders, the "some other race" category, and people who list two or more race groups. These categorizations cover 95% of our core sample and the remainder are coded as having a missing race.

Spouse Race. We assign race and ethnicity to spouses analogously to own race. Because race is an attribute of the individual rather than of their parents, we are able to define race even for spouses that we cannot match to parents. We are able to assign spouse race for 95% of married individuals in our primary analysis sample. When measuring marriage outcomes, we assign those with missing spouse race to an "out-of-sample" race category.

Marriage. Our primary definition comes from tax records at age 30. A person who filed a tax return with the status "married filing jointly" or "married filing separately" is considered married. Anyone who filed a tax return with a different status or does not file a return in the year they are 30 is coded as single. We construct marriage outcomes for each individual in our primary analysis sample and associate married individuals with their spouse's PIK. In addition to our baseline measure, we are able to measure marriage outcomes at age 37 for the 1982 birth cohort, the first in our sample.

Cohabitation. Cohabitation is defined using the set of individuals who received the ACS in the calendar year in which they turn 30. The survey is at the household level and individuals are cohabitating with a partner if they are living with a spouse or un-married partner in the survey.

Childhood Locations. We measure childhood location between child ages 0 and 18 using data on parent locations. For years prior to 2003, we assign parent location from the 1040 form in years where a tax return is filed. Starting in 2003, we use addresses on W-2's and other information returns for non-filers. All addresses are geocoded and assigned to standard Census geographic units (e.g., block, tract, and county) by Census staff in a procedure described by Brummet et al. (2014). In years where no address is assigned, we impute using the chronologically closest non-missing parental address, choosing randomly in the case of ties. Our primary analysis uses childhood location at age 18, which we define using parental location data for all children in our sample, regardless of whether the child is alive and living with their parents at age 18.

Adulthood Locations. After age 18, we define location using an individual's own location rather than their parent(s)' location. If the individual does not file, but an information return is available, that address is used. If not, the location is imputed using the chronologically closest address including the parental address data. At age 27, we are able to assign an address to 99.8% of the sample.

Income in Adulthood. We measure an individual's income using their total pre-tax income when they are 30 years old, which we measure at the household and individual level. If an individual files a tax return, we define household income as the sum of Adjusted Gross Income, social security payments and tax-exempt interest payments on the 1040 tax return. We define individual income as wage income reported on the W-2 plus self-employment and other non-wage income reported on 1040 tax returns. For non-filers, we define both individual and household income as wages from the W-2, or as 0 if no W-2 is filed. We assign percentile ranks relative to other individuals who appear in our child and parent linked data and who are born in the same year.

Sex and Birth Cohort. Sex and birth cohort are obtained from the 2020 Numident file.

Nearest Neighbor Census Tracts. For each census tract, we ordinally rank the nearest 50 census tracts using the geographic distance between census tract centroids.

II.C Summary Statistics

Tables I and II report summary statistics separately by class and race. There are 31.1 million individuals in our analysis sample, all of whom are assigned class by construction, and 95% of whom have non-missing race information. We define class based on parental income quartiles so that 25% of the sample is in each class group.

There are substantial differences in the socioeconomic status of our race and class groups. Individuals with parents in the bottom quartile have a median parent household income of $\$15,010$, compared to $\$136,400$ for individuals with parents in the top quartile. Only 39% of individuals with bottom-quartile parents had married parents in the year they were claimed versus 92% for individuals with parents in the top quartile. White individuals have median parental earnings of $\$72,280$ and 80% have married parents when they are claimed. This compares to $\$29,610$ and 29% for Black individuals.

Consistent with prior research, we find large disparities in adult household income by race and class (e.g., Chetty et al., 2020; Mazumder, 2005; Solon, 1992). Individuals with bottom-quartile parents have a mean household income percentile rank of 37 at age 30 compared to 63 for those with top-quartile parents. The mean rank for white individuals is 55 compared to 36 for Black individuals.[11]

Disparities in household income by race and class are closely tied to differences in marriage rates and spousal incomes (Bayer and Charles, 2018; Binder et al., 2022; Chadwick and Solon, 2002). At age 30, 24% of individuals with bottom-quartile parents are married, compared to 44% for those with top-quartile parents. Among married individuals, those from families in the bottom quartile marry a spouse with an average individual earnings percentile of 52 versus 67 for those from top-quartile families. At age 30, 43% of white individuals and 12% of Black individuals are married. Among those who are married, spouses of white individuals have an average individual earnings percentile of 61 compared to 53 for spouses of Black individuals. White individuals and those from high-income families are more likely to be married and have higher earning spouses than those who come from low-income families or are Black.

To estimate the impact of segregation on these disparities in marriage outcomes, we use sex ratios in childhood neighborhoods as an instrument for marriage market conditions in young adulthood. This relies on being able to measure location for a large fraction of the sample and individuals tending to live, as adults, near where they grew up. We are able to measure location for 87% of our sample at age 27 and 99.8% of the sample after imputing location with the nearest non-missing year with a valid address. At age 27, 67% of individuals in our primary sample still live in their childhood commuting zone, while 56% live in their childhood county, and 51% live within 50 census tracts of their childhood home. Individuals from high-income families move farthest from home, but even among this group, 62% still live in their childhood commuting zone, and 45% live within 50 census tracts of their childhood home.

[11] At the median of the age 30 income distribution, each percentile corresponds to approximately $\$1,300$ in additional household income and $\$1,100$ in additional individual income.

III Descriptive Facts on Marriage, Homophily, and Neighborhoods

III.A Marital Homophily at the National Level

We begin by characterizing overall levels of marital homophily by class in Figure I Panel *A*. At age 30, we find 19.0% of individuals from high-income (top quartile) families have a spouse who also comes from a high-income family. Only 3.1% of those from low-income (bottom quartile) families have a spouse from a high-income family—a difference of 15.9pp. Individuals from low-income families are 1.9pp more likely to marry a spouse from a low-income family than those from high-income families. This smaller gap in the likelihood of marrying a spouse from a low-income family is partly driven by the lower overall marriage rates for individuals from low-income families documented in Table I.

To quantify the extent of homophily, we compare these rates of own-class and interclass marriage to two benchmarks. We show these benchmarks in Figure I Panel *A* and include results for all four class groups in Table III. In the first, which we label the "fully random benchmark," we compute the marriage outcomes that would occur if everybody marries at the same rate and selects a spouse irrespective of class background. The second, which we label the "random spouse benchmark," holds the class-specific marriage rates fixed and computes marriage outcomes under the hypothetical condition that everyone chooses a spouse irrespective of class from the existing pool of married people. The fully random benchmark predicts that 7.4% of people from low- or high-income families would have a spouse from the other class group, compared to a 6.3% prediction by the random spouse benchmark. That is, even if we maintained class differences in marriage rates, eliminating homophily in marriage pairings would substantially increase rates of interclass marriage, closing around 75% of the gap between the factual pairings and those in the fully random benchmark.

Figure I Panel *B* plots the analogous set of statistics for interracial marriage, while Table IV includes marriage outcomes and benchmarks for all race groups. At age 30, only 2.1% of Black individuals have a white spouse, compared to 38.2% of white individuals. Group differences in marriage rates play a larger role for race than class. In the random spouse benchmark, 8.6% of Black people would have a white spouse versus 21.8% in the fully random benchmark, indicating that changing who people marry without changing marriage rates would make up less than half of the gap between observed interracial marriage rates and the fully random benchmark. Even so, eliminating homophily while maintaining racial difference in marriage rates would increase interracial marriage rates by a factor of four.

We next turn to evaluating whether the low rates of interracial and interclass marriage are sensitive to our measure of marriage. In recent decades, Americans have become more likely to marry at a later age and live with a cohabiting partner as an alternative to marriage (Cherlin, 2010; Kennedy and Bumpass, 2008; Lundberg, Pollak and Stearns, 2016; Manning and Cohen, 2012). Measures of homophily will be sensitive to the age at which people marry and type of partnership if those who marry at later ages or opt for cohabitation are more likely to be in interclass or interracial relationships. In Tables A.1 and A.2, we focus on the set of individuals who receive the ACS at age 30 and include married couples and cohabiting partners. In Tables A.3 and A.4, we measure marriage for the 1982 cohort at age 37 using 2019 tax data. In both cases, we find higher overall rates of partnership or marriage than in our primary estimates, but the relative difference between observed rates of interclass and interracial marriage and our benchmarks remains similar to our primary estimates. For example, when looking at cohabitation in Table A.2, the random spouse benchmark implies a share of Black individuals with a white spouse that is 4.6 times higher than the factual pairings, compared to 4.1 times higher in our primary estimates.

Rates of interclass and interracial marriage are low for both sexes. Tables A.5 and A.6 show marriage outcomes by class and race separately by sex. Females are more likely to be married at age 30 than males. Among married individuals from low-income families, 13.4% of men and 12.6% of females have a spouse from a high-income family. There are slightly larger differences by sex for interracial marriage (Fryer, 2007). Among married Black people, 22.2% of males and 14.7% of females have a white spouse. Among all Black people, 2.6% of males and 1.7% of females have a white spouse. We focus on estimates pooling both sexes for the remainder of our empirical analyses.

III.B Neighbors as Possible Marriage Partners

The impact of segregation on interclass and interracial marriage depends on whether who lives nearby influences who one marries. Potential spouses can meet within their neighborhoods, and these communities can shape one's professional and social circles (Athey et al., 2021; Chetty et al., 2022; McPherson, Smith-Lovin and Cook, 2001). For example, online dating platforms have become increasingly popular in recent years (Rosenfeld and Thomas, 2012; Rosenfeld, Thomas and Hausen, 2019), and a recent survey shows that 35% of online daters are only willing to date within their neighborhood (Kirkham, 2019). In this section, we document the extent to which people tend to marry somebody who had lived near them the past.

We begin with a definition of neighborhood that includes an individual's home census tract and the 50 nearest census tracts. Census tracts are constructed to represent similar populations, meaning our definition will span a larger physical area in rural towns, and a smaller area in cities. In 2010, there were 74,134 census tracts in the U.S. Figure A.1 shows what this definition looks like in areas with different population densities. The nearest 50 census tracts to Harvard University, for example, exclude almost the entirety of the city of Boston. For a more rural area like Gilford, New Hampshire, the nearest 50 census tracts include the majority of central New Hampshire. In general, the average share of the U.S. population that lives within 50 census tracts of a given person is 0.07%. If people chose a spouse irrespective of distance or if marriage markets were substantially larger than 50 census tracts, a small share of marriages will be between two people who have lived within 50 census tracts of each other.

Even though the nearest 50 census tracts encompasses a small share of the overall population, we find that a large fraction of marriages are between two people who had lived within 50 census tracts of each other prior to marriage. In Figure II Panel *A*, we show the fraction of couples who lived within 50 census tracts of each other by the year relative to marriage. We find that five years prior to marriage, 68% of couples had already lived within 50 census tracts of each other at some point in their lives. This is partly due to the fact that couples have already moved in with each other. When we remove these observations, we find that 58% of couples have already lived within 50 census tracts of each other, but not at the same address, five years prior to their marriage.

Even after excluding couples at the same address, these rates can still be driven by couples who choose to live near each other, but not at the same address, in the years leading up to their marriage. Ten years prior to marriage—a time when 90% of eventually married couples have yet to meet each other (Rosenfeld, 2017)—we find that 45% of those couples have already lived within 50 census tracts of each other (excluding those at the same address). We show these statistics for a variety of geographic definitions, ages and years, relative to marriage in Table V. Even going back to birth, we find that 23% of couples are born within 50 census tracts of each other. Across ages, the share of couples living within 50 census tracts is similar to the share who are living within 10 miles of each other or in the same county.

Within 50 census tracts, individuals are more likely to marry somebody in the nearer census tracts. Figure II Panel *B* plots marriage probabilities relative to somebody in one's own census tract based on neighborhood at age 18. Individuals are nearly 20 times more likely to marry somebody from their own tract than somebody who lives 50 tracts away. The probability of marrying somebody within one's own tract or the nearest nine tracts is equal to the probability of marrying somebody between 10 and 50 tracts away.

III.C Segregation by Race and Class in Childhood and Adulthood

Having shown that people tend to marry somebody from their neighborhood, we turn to documenting neighborhood segregation by race and class. We measure neighborhood composition by race and class in the nearest 50 census tracts and nine birth cohorts (own cohort and the four older/younger cohorts). Within 50 tracts, we weight individuals based on their geographic and age distance.[12] For geographic distance, we weight individuals based on how likely they are to marry relative to somebody in their own tract based on the pattern shown in Figure II Panel *B*. For example, a person living 25 tracts away will receive a weight of 0.11, since the probability of marrying somebody 25 tracts away is $\frac{1}{0.11} \approx 9.1$ times smaller than the probability of marrying somebody from one's own tract. In practice, we allow the weights to vary by terciles of population density since marriages are more spatially concentrated in rural areas—the weights by population density are shown in Figure A.2 Panel *A*.

For age, we use the weights in Figure A.2 Panel *B* and allow the weights to vary by sex, since men are more likely to marry somebody who is younger and women are more likely to marry somebody who is older. We fix the weight of the modal age difference to be one. For men, the birth cohort one year younger gets a weight of 0.97, since men are almost equally likely to marry somebody one year younger as they are to marry somebody from their own cohort, the modal age difference for men. For women, the birth cohort one year younger has a weight of 0.52 since women are $\frac{1}{0.52} \approx 1.9$ times more likely to marry somebody one year older than one year younger, the modal age difference for women.[13]

For the remainder of the paper, we use the phrase "neighbors" to refer to the group of people living within 50 census tracts, according to the weights described above. For example, the "fraction of neighbors who are Black" is the fraction of those in the nearest 50 tracts and nine birth cohorts who are Black, weighted by the distance and relative age weights. In practice, this means that most of the weight is on those living within 10 tracts and who are between 0 and 3 years younger for men and 0 to 3 years older for women.

Figure III shows the distribution, across people in our sample, of the fraction of neighbors who come from a given race or class group. We focus on the neighborhoods people live in during young adulthood, which we define as the ten-year period covering ages 18 through 27. Each age receives an equal weight. Panels *A* and *B* show segregation by class, while panels *C* and *D* show segregation by race.

In young adulthood, the median person from a high-income family lives in a neighborhood where 34% of their neighbors also come from a high-income family. For the median person from a low-income family, only 17% of their neighbors come from high-income families, while the median person from a high-income family has twice as many neighbors from high-income families than people from low-income families. Similarly, people from low-income families have 1.6 times more neighbors from low-income families than people from high-income families.

We find similar differences by race. In young adulthood, the median white person lives in a

[12] The construction of the weights is described in more detail in Appendix A.

[13] The modal age difference need not be the same for men and women since it is possible to marry a spouse who is not in our primary sample.

neighborhood where 78% of their neighbors are white. For the median Black person, only 45% of neighbors are white. White people have 1.7 times more white neighbors than Black people do. The median Black person lives in an environment where 38.2% of their neighbors are Black, which is 5.6 times higher than for the median white person.

III.D Association Between Segregation and Marriage Outcomes

Having shown that people tend to marry neighbors and that neighborhoods are segregated, we now shift our focus to examining the association between segregation and marriage outcomes. Despite high average levels of segregation by race and class, Figure III shows substantial variation across people in neighborhood exposure to other race and class groups. In this section, we use that variation to examine whether people with more neighbors from other race and class backgrounds are more likely to have an interracial or interclass marriage.

We begin by focusing on individuals from low-income families. In Figure IV Panel *A*, we produce a binned scatter plot of interclass marriage versus segregation. We split the sample into ventiles based on the fraction of neighbors between ages 18 and 27 who come from high-income families. Within each ventile, we plot the average percentage of individuals who have a spouse from a high-income family against the average fraction of neighbors from high-income families. The dashed line, which we label the "random sampling benchmark," has a slope equal to the marriage rate (and an intercept equal to zero) among individuals from low-income families and therefore represents what this relationship would look like if people choose a spouse from their pool of neighbors regardless of class background.

The series forms a clear upward sloping relationship. Among individuals from low-income families, a 10pp increase in the fraction of neighbors who come from high-income families leads to a 2.2pp increase in the probability of having a spouse from a high-income family. Only 0.3% of the individuals in the bottom ventile of exposure to high-income families (4.4% of neighbors) have a spouse from a high-income family, compared to 8.5% of the individuals in the top ventile of exposure to high-income families (45.5% of neighbors). The slope of 0.22 also closely aligns with the random sampling benchmark of 0.24, indicating the association between marriage outcomes and exposure for individuals from low-income families is similar to what we would anticipate if people select spouses from their pool of neighbors regardless of class.

In Figure IV Panel *B*, we build an identical plot for individuals from high-income families and focus on exposure and marriage to individuals from low-income families. We again see a clear upward sloping relationship with a slope of 0.14, although the slope slightly flatter than the 0.22 slope for individuals from low-income families. Nonetheless, individuals from high-income families are substantially more likely to have a spouse from a low-income family if their neighborhood in young adulthood has a higher fraction of their neighbors are from low-income families.

The relationship between intergroup marriage and cross-group exposure is flatter for race than class. Figure IV Panels *C* and *D* repeat our binned scatter plots by race instead of class. The slope for white individuals is 0.02—a 10pp increase in the share of their neighbors who are Black is associated with a 0.2pp increase in the probability of having a Black spouse. For Black individuals, we obtain a slope of 0.09, which implies a 0.9pp increase in the probability of having a white spouse when the fraction of neighbors who are white increases by 10pp. This relationship is driven primarily by Black individuals living in majority white neighborhoods—there is a substantially weaker association in neighborhoods with fewer white residents.

We combine the panels of Figure IV to summarize these relationships by race and class in Figure V. In Panel *A*, we form a single sample containing individuals from high- and low-income families and produce a single binned scatter plot. To do so, we create ventiles based on the fraction of neighbors who are from the "other" class group, meaning that for those from high-income families the ventiles are based on the fraction of neighbors who come from low-income families. The series in Panel *A* plots the likelihood of having a spouse from the other class group against neighborhood exposure to the other class group. Panel *B* is an analogous plot for race using a combined sample of white and Black individuals. We use this combined sample for the majority of our remaining empirical analyses.

IV Quasi-Experimental Effects of Neighborhood Exposure

The association between segregation and marriage outcomes documented in Section III.D compares people who choose to live, as adults, in demographically different neighborhoods. The fact that people are more likely to have an interracial or interclass marriage when they live in neighborhoods with more people of other race and class backgrounds could represent a causal effect of exposure, selection, or even reverse causality. People could choose to live in neighborhoods with more members of a given race or class group if they have traits that make them more appealing to members of that group or if they have a preference to marry somebody from that group. People could also be living in neighborhoods with more members of a given race or class group because they are already in a relationship with somebody from that group. In this section, we develop an instrument for exposure to opposite-sex members of other race and class groups and estimate the causal effect of exposure on interracial and interclass marriage.

IV.A Sex Ratios in Childhood Neighborhoods

We use sex ratio variation in childhood neighborhoods as an instrument for exposure to opposite-sex members of other race and class groups. Our approach leverages random sex assignment at birth and is related to other strategies that use quasi-experimental variation to generate exogenous changes in sex ratios (Abramitzky, Delavande and Vasconcelos, 2011; Angrist, 2002; Bazzi et al., 2023; Brainerd, 2007; Charles and Luoh, 2010; Zaremba, 2022).

The instrument impacts market "tightness," which we define as the difference in the fraction of opposite versus own-sex neighbors who come from a given race or class group. For example, a white male born in a year and place where nearby Black families tended to have more female than male children faces a "tighter" interracial marriage market. Segregation is fundamentally about market "thickness," the overall fraction of neighbors who come from a given race or class group. In Section V, we build a spatial model of the marriage market that takes our IV estimates of market tightness as inputs and predicts what would happen in response to changes in market thickness, i.e. segregation. In this section, we focus on estimating the causal effect of changes in market tightness on interclass and interracial marriage.

Our approach relies on three facts. The first is that, as of 2020, 99% of legal marriages in the U.S. were opposite-sex marriages (Kreider et al., 2023). We measure marriage from 2012-2019 (same-sex marriage was illegal in many states until the Obergefell v. Hodges Supreme Court decision in 2015). Because legal marriage—which we can measure on tax returns—is almost exclusively an opposite-sex phenomenon during our sample period, we use sex ratio variation and define market tightness from the perspective of opposite-sex relationships.

The second fact is that individuals tend to live, as adults, near their childhood neighborhoods (Foster, 2017; Sprung-Keyser, Hendren and Porter, 2022). We show in Tables I and II that the majority of Americans live within 50 census tracts of their childhood home at age 27. As a result, the demographic composition of childhood neighborhoods impacts the demographic composition of adulthood marriage markets.

The third is that it is common to marry a nearby neighbor who is close in age. We showed in Section III.B that people tend to marry somebody who lived nearby in the past. As a result, neighborhood sex ratios in nearby birth cohorts impact the demographic composition of an important component of the marriage market but are also small enough to generate sex ratio variation that we can use for IV estimates.

Our baseline estimates measure childhood neighborhood using parent locations at child age 18. If parent location at age 18 is missing because parents did not receive tax or information returns in that year, we impute location using the closest age less than 18 with non-missing data. We assign a childhood neighborhood even if the child has already died before age 18 or is no longer living with their parents. We are able to assign a childhood neighborhood to more than 99% of the sample.

One helpful feature of our approach is that the data generating process of the instrument—sex assignment via a Bernoulli trial—is already known. As a result, we can non-parametrically test whether the observed distribution of our instrument is consistent with the distribution we would expect to see if sex is randomly assigned to each individual in the data. Our instrument would depart from Bernoulli variation if parents sort into neighborhoods based on the sex of their child.

In Figure VI we compare the observed distribution of the instrument to the one that would emerge under random Bernoulli sex assignment. To do so, we draw from a binomial distribution at the census tract and group level for each of our four key groups: individuals from low/high-income families and white/Black individuals. This is equivalent to randomly assigning sex to each individual in the data and then computing the sex ratio for each tract and group. We set the success probability of the binomial equal to the national fraction of children who are female and run the simulation 1,000 times. We estimate each percentile of the distribution by taking the mean estimate of that percentile across simulations. The observed distribution of the instrument closely matches the distribution that we would expect to see if the variation were coming exclusively from random sex assignment.[14] Consistent with this finding, Figures A.3 and A.4 show that the instrument is uncorrelated within a tract across groups and unrelated to tract by birth cohort covariates, such as mean parental income.

IV.B Market Tightness, Interracial and Interclass Marriage

We define our endogenous treatment variable,

$$T_{i,g} = \frac{N^a_{i,-s,g}}{N^a_{i,-s}} - \frac{N^a_{i,s,g}}{N^a_{i,s}},$$

for person i and race or class group g is the difference, in adult census tracts from ages 18 to 27, between the share of opposite- and own-sex neighbors who belong to group g. Neighbors include individuals within 50 census tracts and four birth cohorts. The variable $N^a_{i,s,g}$ is the weighted count of neighbors of person i, between ages 18 and 27, who are of sex s and group g. $N^a_{i,s}$ is the (weighted) total number of neighbors of person i who are sex s. The $-s$ indicates counts for the opposite sex and s is used for own sex. The weights, based on census tract distance and relative age, are described in detail in Section III.C.

[14]We can see in VI that the instrument even matches the integer breaks that come from repeated integer values in some cells.

Our instrument,

$$Z_{i,g} = \frac{N^c_{i,-s,g}}{N^c_{i,g}},$$

is the fraction of group g children in person i's childhood neighborhood that belong to the opposite sex. The variable $N^c_{i,-s,g}$ is the weighted count of the number of opposite-sex members of group g in person i's childhood neighborhood. The variable $N^c_{i,g}$ is the total (weighted) count of group g people in i's childhood neighborhood.

We use the following specification for our first stage and reduced form:

$$T_{i,g} = \beta^f Z_{i,g} + \delta^f_{o,l,s} + \Gamma^f X_{o,l,s,g} + \varepsilon^f_{i,g} \quad (1)$$

$$Y_{i,g} = \beta^r Z_{i,g} + \delta^r_{o,l,s} + \Gamma^r X_{o,l,s,g} + \varepsilon^r_{i,g}. \quad (2)$$

We instrument for market tightness, $T_{i,g}$, in the marriage market for group g with sex ratios among group g in person i's childhood neighborhood $Z_{i,g}$. Our first stage coefficient is β^f. The o subscript indicates own group—in specifications focused on class, o indicates own-parent quartile, and for race the o indicates own race group. We include own group by childhood census tract (l) by sex fixed effects, $\delta_{o,l,s}$, so that we are never comparing across neighborhoods or groups. The within childhood tract variation compares two children of the same sex and group faced different sex ratios among group g because they were born in different years.

If $T_{i,g}$ and $Z_{i,g}$ were measured contemporaneously (rather than one in adulthood and one in childhood), they would be related to each other by Bayes' rule. We show in Appendix B that, conditional on the marginal probabilities that appear in Bayes' rule, $T_{i,g}$ is a linear function of $Z_{i,g}$. Because of this relationship, we control for the marginal probability expressions in $X_{o,c,s,g}$, which includes own-group and sex fixed effects interacted with the overall share of (weighted) people who belong to group g and the overall share who belong to sex s in person i's childhood neighborhood.

The outcome variable, $Y_{i,g}$, is an indicator equal to one if person i is married to a member of group g at age 30, and otherwise equals zero. We will also estimate effects on the overall likelihood of being married. Specifications by class use a sample of individuals from families in the bottom and top quartile of parent household income, and for these results g is the "other" class group. For example, an individual from a top-quartile family will have $Y_{i,g}$ equal to one if and only if they are married to a spouse from a bottom-quartile family (vice versa for those from bottom quartile families). For specifications by race, white individuals have $Y_{i,g}$ equal to one if they have a Black spouse, while Black individuals with a white spouse have $Y_{i,g}$ equal to one.

The distribution of childhood neighborhood sex ratio variation based on our definition using the nearest 50 census tracts and four birth cohorts is shown in Figure VII Panels *A* and *C*. In Panel *A*, we use a sample of individuals from families in the bottom and top parent income quartiles and focus on sex ratios in the other class group. For those from the top quartile, the instrument is the fraction of bottom quartile individuals in their childhood neighborhood who are the opposite sex (and vice versa for individuals from families in the bottom quartile). This histogram shows the residual distribution of the instrument in equation 1. In Panel *A*, nearly all of the variation is between $[-3,3]$. For example, in a neighborhood with 500 members of the other class group, the instrument can change the environment from one with 235 opposite-sex and 265 own-sex individuals, to one with 265 opposite-sex and 235 own-sex individuals (a "shortage" vs. "surplus" of 30 people).

In Panel *C*, we can see that our instrument has slightly more residual variation when we look by race rather than class. This is because the typical number of neighborhood residents of the other race group—i.e. Black neighbors for white individuals and white neighbors for Black individuals—is smaller than the typical number of residents of the other class group. The instrument is almost entirely contained in $[-5,5]$. This suggests that for a white individual in a neighborhood with 500 Black residents, the instrument can change the environment from one with 225 opposite-sex and 275 own-sex Black individuals, to one with 275 opposite-sex and 225 own-sex Black individuals (a shortage vs. surplus of 50 people).

We overlay the first stage relationship between childhood neighborhood sex ratios and adult marriage market tightness in Figure VII Panels *A* and *C*. We follow Cattaneo et al. (2019) and modify equation 1 to regress $T_{i,g}$ on indicators for each ventile bin of $Z_{i,g}$. We plot the coefficients on each ventile bin against the mean value of $Z_{i,g}$ in that ventile bin and normalize the mean of the coefficients and the mean of $Z_{i,g}$ to zero. This method allows us to non-parametrically evaluate the first stage relationships before proceeding with the linear specification in equation 1. We report the coefficient from the linear model in equation 1 in Panels *A* and *C*.

We see a strong first stage relationship between sex ratios in childhood neighborhoods and marriage market tightness in adulthood. The first stage for class is steeper than race, which reflects the fact that the overall share of neighbors from the other class group is larger than the overall share of neighbors from the other race group. In either case, individuals born in places and years where a larger fraction of children from other race or class groups were of the opposite sex tend to live as adults in neighborhoods where more of their opposite-sex neighbors come from other race or class groups relative to their own-sex neighbors.

We plot the reduced form relationship between interclass and interracial marriage and market tightness in Panels *B* and *D* of Figure VII. Using methods analogous to Panels *A* and *C*, we plot the reduced form relationship using the Cattaneo et al. (2019) method and report the coefficient on β^r from equation 2.

In Panel *B*, we see a clear and significant upward sloping relationship between interclass marriage and exposure across class lines. Individuals growing up in places and years where children of the other-class group were more likely to be opposite sex are more likely to be married to a spouse from the other class group as adults. We report IV coefficients for the pooled sample as well as separate estimates for individuals from low- and high-income families in Figure VIII. The pooled sample coefficient of 0.13 implies that going from a marriage market where, for example, 20% of the own- and opposite-sex neighbors are from the other class group to one where 15% of the own-sex and 25% of the opposite-sex neighbors are from the other class group would generate a 1.3pp increase in the likelihood of having a spouse from the other class group. The IV estimates are similar for individuals from low- and high-income families.

Unlike class, Panel *D* shows that exposure across race lines has no detectable impact on interracial marriage. White and Black people growing up in years and places where children of the other race were more likely to be opposite sex are no more likely to be in an interracial marriage as an adult. Figure VIII shows IV estimates separately for white, Black and the pooled sample. Although we are unable to rule out small effects of exposure on interracial marriage, we are able to clearly reject exposure effects of the same magnitude as the class estimates.

In addition to showing that the distribution of our instrument is consistent with random Bernoulli sex assignment, Table VI shows that our IV estimates are similar when we use within family variation in market tightness, ruling out any possible role of parental sorting. In column 3, we use variation in market tightness between siblings. This variation comes both from the years each sibling was born and from their childhood neighborhoods. In column 4, we interact the family fixed effects with childhood census tract to use variation exclusively from birth timing among siblings who were assigned the same childhood neighborhood because their parents did not move. In both cases, we get estimates similar to our baseline estimates in column 2.

Market tightness influences interclass marriage by shaping who people marry, rather than by affecting their chances of getting married in the first place. Column 5 of Table VI shows that market tightness has a minimal and statistically non-significant impact on the likelihood of being married. Consistent with this evidence, Figure A.6 shows that market tightness matters in a class-specific manner. For example, market tightness among neighbors from bottom parent income quartile families has the largest impact on the likelihood of marrying a spouse from a bottom parent income quartile family. Market tightness among the other quartiles have small, negative, and insignificant, effects on the likelihood of having a spouse from a bottom quartile family.

Our baseline estimates define childhood neighborhoods using parent locations at child age 18. Choosing a later age helps improve the precision of the IV estimates because it brings the measurement of the childhood sex ratios chronologically closest to our adult exposure estimates, and even includes a year of overlap since adult exposure is measured between ages 18 and 27. To ensure our definition of childhood neighborhood at age 18 is not driving our results, we estimate our IV model using each age from 0 to 18 to define childhood neighborhoods and measure the instrument. Figure A.5 shows that our baseline result of a large exposure effect by class and small or zero effect by race holds even going back to the neighborhoods where individuals were born. Variation in sex ratios in the neighborhoods individuals live in as adults, rather than those where they grew up, can generate biased IV estimates if individuals sort to neighborhoods on the basis of sex (e.g., certain areas have jobs with a high fraction of male workers). In Figure A.5 we see that the IV estimates quickly increase once we use variation in adult neighborhoods.[15]

How can we be sure that sex ratios in childhood neighborhoods impact marriage outcomes through their effect on adulthood marriage markets? Perhaps going to school with opposite-sex members of a given race or class group could impact marriage outcomes by, for example, learning how to interact with opposite-sex members of a given group. If this were driving our results, we might see that having more opposite-sex children of other class groups in one's own birth cohort and neighborhood would increase the likelihood of having an interclass marriage to an individual outside of that cohort or neighborhood range.

In Figure IX, we evaluate the impact of cohort-specific sex ratios on cohort-specific marriage outcomes. In Panel *A*, we show how cohort-specific sex ratios among one's own class impacts the probability of having an own-class spouse who was born in the prior cohort. We repeat this analysis in Panel *C* with other-class sex ratios and interclass marriage, and repeat both plots in Panels *B* and *D* using marriage to a spouse born in the cohort after one's own. Across all four panels, we see that the sex ratio with the largest impact on marriage outcomes is the one in the cohort for which marriage outcomes are measured. Having more people of the opposite sex in the other class group of one's own cohort does not impact the likelihood of having an interclass marriage in the preceding or following birth cohort. We also show in Figure A.7 that our IV estimates tend to generate marriages between two people who shared the same childhood neighborhood, consistent with exposure rather than learning driving the results.

The childhood sex ratio instrument generates changes in adult marriage market tightness for different race and class groups. Tighter marriage markets across class lines generate an increase in interclass marriage, but tighter marriage markets across race lines have no detectable impact on interracial marriage. Tightening a marriage market leaves the overall share of each race and class group in the market fixed, but changes the fraction who belong to each sex. Reducing segregation, therefore, requires changing market thickness, the overall share of each race and class group in a neighborhood.

[15] After age 18, neighborhoods are assigned using an individual's own data rather than their parents.

IV.C From Market Tightness to Market Thickness

There are two key limitations of our IV estimates for studying the changes in market thickness that would be brought about by changes in exposure through changes in residential segregation. The first, which is briefly discussed above, is that market tightness and market thickness are different. Our IV estimates alter market tightness by subbing out a male for a female (or vice versa) who would have otherwise existed only under a counterfactual sex assignment at birth. Reducing segregation changes market thickness by moving people across neighborhood boundaries. This is a different type of variation because it will typically hold sex ratios fixed while changing the overall share of a given race or class group in a neighborhood.

In addition to effects in the origin and destination neighborhoods, there can be spillover effects across neighborhood boundaries. Because marriages can occur across neighborhood boundaries, any counterfactual change to marriage outcomes of residents of one neighborhood will necessarily affect residents of others. Our IV results estimate the effect of market tightness on marriage outcomes for residents of a single neighborhood, without considering possible spillovers across neighborhood lines.

We visualize this distinction with a simple three-neighborhood example in Figure X. In this example, there are two types of people, low- and high-income, and we consider the marriage market impact of moving a low-income resident of neighborhood C to neighborhood A. The solid lines show the initial marriage outcomes and the dashed lines indicate the counterfactual marriage outcomes that would occur after the move. In Panel *A*, the person labeled $\star$ marries the person labeled 1, and in Panel *B*, $\star$ marries person 2. The move generates one new interclass marriage in Panel *A* and two in Panel *B*. Yet, from the perspective of the treated neighborhood, A, the impact of the move on interclass marriage was one in both panels.

Our goal is to estimate the impact of reducing segregation on rates of interclass and interracial marriage. To do so, we focus on market thickness, rather than market tightness, and account for the total (general equilibrium) effect on the marriage market rather than the partial equilibrium effect in a single neighborhood. In the next section, we develop and estimate a spatial model of the marriage market that allows us to perform this task.

V A Spatial Model of the Marriage Market

We use our model to quantify the impact of segregation on interclass and interracial marriage. Importantly, we estimate the model using the partial equilibrium moments that we focused on in Sections III and IV, but we then use the model to simulate the impact of segregation in general equilibrium.

We focus on two types of counterfactuals. First, we simulate the impact of programs that reduce segregation by moving a small share of a city's overall population. These programs either help individuals living in high-poverty areas access low-poverty neighborhoods, such as Moving to Opportunity or the Gautreaux Project (e.g., Bergman et al., 2019; Chyn, Collinson and Sandler, 2023; Ludwig et al., 2013), or bring middle-income residents to formerly high-poverty neighborhoods, such as the Hope VI program (e.g., Haltiwanger et al., 2020; Popkin, 2004). Second, we use the model to estimate what rates of interclass and interracial marriage would look like absent segregation. This exercise quantifies the total impact of segregation on marital homophily, approximating the impact of policies such as land use restrictions, which have been shown to have large impacts on segregation (e.g., Glaeser and Gyourko, 2002; Rothwell and Massey, 2009, 2010).

V.A Model Setup

We start from the Choo and Siow (2006) transferable utility model of the marriage market. There are two sides to the market, which we refer to as "men" and "women." Each individual belongs to one of T possible types; there are m_j total type j men in the market and w_k type k women. Individuals have preferences over potential spouses that have two components: a "systematic" component that all type k women have over type j men (and an analogous component for type j men over type k women) and an idiosyncratic component that varies at the individual, by spouse-type level. Market clearing transfers, τ_{jk}, are set so that the number of type k women who want to marry a type j man is equal to the number of j men who want to marry a type k woman.[16] To bring the Choo and Siow (2006) model to our setting, we depart from prior applications in two important ways.

First, we explicitly model the role that space plays in the marriage market by including census tract in the type space and allowing individuals to have preferences over distance from their spouse. This contrasts with prior applications that treat the U.S. as one large marriage market or use coarse regional groupings such as states (e.g., Chiappori, Oreffice and Quintana-Domeque, 2012, 2018; Mourifié and Siow, 2021; Siow, 2015). The large market assumption is rejected in our setting because individuals disproportionately marry their neighbors, and hyperlocal changes to neighborhood composition impact marriage outcomes. If individuals prefer to marry somebody who lives near them, rates of interclass and interracial marriage can arise from a combination of segregation—neighbors are more likely to come from the same race and class group—and a preference to marry within group. Including neighborhood in the type space allows preferences and segregation to play a role, but introduces computational challenges which we address in Section V.B.

[16] The transfers need not be monetary transfers but can include implicit, non-monetary, transfers between partners.

Our second departure is to allow for the possibility that there is a trait in the type space that is unobserved to the researcher, but is correlated with the observed traits. We refer to this trait as the "unobserved trait." Without this type of econometric endogeneity, any association between neighborhood exposure and marriage outcomes is assumed to be causal. However, we showed that the association between cross-race exposure and interracial marriage is positive (Figure V), even though our IV effect of cross-race exposure and interracial marriage is near zero (Figure VIII).

The presence of the unobserved trait allows us to account for sorting into neighborhoods. For example, Black individuals who live in majority white neighborhoods during adulthood might have characteristics that are unobserved to the researcher that lead them to prefer or be preferred by a potential white spouse. In this scenario, the observed association between neighborhood white exposure and interracial marriage rates would not represent the causal effect of exposure. The data moments we use to estimate the model include our IV estimates. Our instrument, random sex ratio assignment at birth, is orthogonal to any neighborhood sorting on the unobserved trait.

The types in the model, men indexed by j and women by k, are defined at the intersection of

$$j,k \in \{\text{Parent Income Quartile} \times \text{Race} \times \text{Adult Census Tract} \times \text{Unobserved Trait}\}.$$

We model the unobserved trait as a binary characteristic. Preferences for men,

$$U^m_{k,i} = \alpha^m_{j(i),k} - \tau_{j(i),k} + \varepsilon_{k,i}, \tag{3}$$

and women,

$$U^w_{j,i} = \alpha^w_{j,k(i)} + \tau_{j,k(i)} + \varepsilon_{j,i}, \tag{4}$$

are a function of the systematic component, the transfer, and the idiosyncratic component. The idiosyncratic component follows a type one extreme value distribution,

$$\varepsilon_{t,i} \sim \text{Gumbel}(0,1). \tag{5}$$

Here, $U^m_{k,i}$ is the preference that person i, who is a type $j(i)$ man, has over marrying a type k woman and $U^w_{j,i}$ is the preference that person i, who is a type $k(i)$ woman, has over marrying a type j man.

If individuals sort to neighborhoods based on whether they have the unobserved trait, failing to model the unobserved trait would imply preferences with unrealistic logit substitution.[17] One

[17] If the unobserved trait is more prevalent in some demographic groups, segregation, which effectively increases the cost of marrying somebody from a given group, could generate substitution patterns inconsistent with the basic multinomial logit. If people with the unobserved trait want to marry each other, then reducing exposure to a particular group would disproportionately increase the probability of marriage to another demographic group where the

way to solve this issue would be to posit a more general distribution for unobserved preference heterogeneity (e.g., Galichon and Salanié, 2022). We do this by allowing our utility functions to include preferences for the unobserved trait. However, rather than assuming the parameters of the more general distribution are known, we estimate the parameters that govern preferences over the unobserved trait. Our approach can be used in other matching applications in which some of the relevant traits are not observed, are potentially correlated with the observed traits, and there is a plausible instrument.

Marital surplus,

$$\gamma_{j,k} \equiv (\alpha^m_{j,k} + \alpha^w_{j,k}) - (\alpha^m_{j,0} + \alpha^w_{0,k})$$

is the value of a j,k marriage net of the utility that j and k would attain if they were single. Choo and Siow (2006) show that the equilibrium number of matches between types j and k,

$$\mu_{j,k} = \sqrt{e^{\gamma_{j,k}} \mu_{j,0} \mu_{0,k}}, \tag{6}$$

is a function of the marital surplus and the number of types j and k who remain single, $\mu_{j,0}$ and $\mu_{0,k}$. The equilibrium also must obey adding up: the total number of type t in the market must be equal to the sum of the number of t who marry and the number who remain single:

$$m_j = \mu_{j,0} + \sum_{k=1}^{T} \mu_{j,k}$$
$$w_k = \mu_{0,k} + \sum_{j=1}^{T} \mu_{j,k}.$$

We can plug the expression for the equilibrium number of matches, $\mu_{j,k}$, into the adding up constraints for a concise definition of the model equilibrium,

$$m_j = \mu_{j,0} + \sum_{k=1}^{T} \sqrt{e^{\gamma_{j,k}} \mu_{j,0} \mu_{0,k}} \tag{7}$$
$$w_k = \mu_{0,k} + \sum_{j=1}^{T} \sqrt{e^{\gamma_{j,k}} \mu_{j,0} \mu_{0,k}}. \tag{8}$$

This is a system of $2T$ equations and $2T$ endogenous variables, the number of each type who choose to remain single.

unobserved trait is prevalent.

We model marital surplus,

$$\gamma_{j,k} = f(j,k;\zeta) + \chi_{c(j),c(k)} + \rho_{r(j),r(k)} + \mathbb{1}\{u(k)=1\}\upsilon_{c(j),r(j)} + \mathbb{1}\{u(j)=1\}\upsilon_{c(k),r(k)}, \tag{9}$$

as a function of the distance between the spouses and the preferences that each person has over the unobserved trait, class, and race. The function $f(.)$ is a continuous function of the census tract distance between a type j and a type k. There is a premium for living within 50 census tracts of each other and then a smooth decay between 0 and 50 census tracts. The ζ parameters govern the within 50 tract premium and the smoothness of the decay function. We describe how $f(.)$ is parameterized in Appendix D.

Match value based on class, $\chi_{c(j),c(k)}$, is a function of the parent income quartile of type j, $c(j)$, and k, $c(k)$ (e.g., there is a surplus for a top-bottom quartile match, bottom-bottom, bottom-second, etc.). The $\rho_{r(j),r(k)}$ is the analogous component for race, where $r(t)$ is a function that gives the race group of type t. If k has the unobserved trait, the surplus includes that value that a person of race $r(j)$ and class $c(j)$ places on the unobserved trait, $\upsilon_{c(j),r(j)}$. If j has the unobserved trait, then surplus includes the type k's value of the unobserved trait, $\upsilon_{c(k),r(k)}$.

The parameters of the model have two components. The first is the marital surplus, $\gamma_{j,k}$. The second component governs the role of the unobserved trait. Each race and class group has two parameters: one determines the overall share of that group that has the unobserved trait and the second governs sorting, or how likely individuals of that group are to have the unobserved trait based on the overall income level of the neighborhood that they live in. We discuss this procedure in detail in Appendix C.

V.B Estimation

With data on equilibrium matches and counts of each type, $\{\mu, m, w\}$, the model is just identified and equations 7 and 8 can be inverted to solve for γ. However, these data are not observed in our setting because of the unobserved trait. We do not observe the number of men of each type, m_j, women of each type, w_k, nor the number of matches between a j and k, $\mu_{j,k}$. Instead, we estimate the model using simulated method of moments (SMM). The SMM objective function minimizes the distance between the model moments and data moments subject to a weighting matrix.

$$\hat{\theta}(W) = \arg\min_{\theta} [\hat{\psi}_d - \hat{\psi}_s(\theta)]^\top W [\hat{\psi}_d - \hat{\psi}_s(\theta)] \tag{10}$$

We use three different types of moments in ψ. The first set of moments are the national marriage outcomes by race and class. For each parent income quartile, we compute the fraction of individuals married to a spouse from each of the other quartiles.[18] We compute the analogous marriage outcomes by race, but aggregate our data into four race groups for simplicity: non-Hispanic white, non-Hispanic Black, Hispanic, and other. National marriage outcomes by class and race are in Tables III and IV.

Our second set of moments are the descriptive associations between marriage outcomes and segregation that we documented in Section III.D. Similar to the national marriage outcomes, for each parent income quartile, we estimate the the association between the probability of marrying somebody from a given parent income quartile and the fraction of neighbors between ages 18 and 27 who come from that same parent income quartile. We construct the analogous estimates by race. Figure IV shows these moments for the four key race and class groups we focus our analysis on.

Our third and final set of moments are the IV estimates of the causal effect of market tightness, i.e. exposure to opposite-sex vs. own-sex members of other race and class groups, on interclass and interracial marriage. For each parent income quartile, we include the IV effects of exposure on marriage to somebody from each of the other quartiles and construct the analogous estimates by race.[19] We include random sex ratio variation in our sample, which allows us to estimate a model analogue of our IV coefficients. Section IV details the construction of the empirical version of these estimates and the IV coefficients for our four key race and class groups are summarized in Figure VIII.

Conditional on θ, γ and {m, w} are known because θ includes the parameters that govern the unobserved trait. The next step is to compute the equilibrium matches by solving the system of equations defined in equations 7 and 8 and plugging the results into equation 6. Unfortunately, this is a computationally costly problem that needs to be solved for each iteration of the SMM objective function as θ changes. The dimension of μ is T^2+2T. With four parent income quartile and race groups plus the dichotomous unobserved trait, $T^2 = 1,024 \times C^2$, where C is the number of census tracts in the sample. The problem quickly becomes computationally infeasible as C gets large.

We deal with this issue in two ways. The first, which we document in Appendix D, is to implement a change of variables in equations 7 and 8 that allow them to be expressed as a system of quadratic equations. This redefined system can be solved using fixed point iterations which is substantially computationally cheaper than non-linear solvers that rely on the $2T \times 2T$ Jacobian matrix. We solve the equations using the technique outlined in Anderson (1965) implemented

[18]The sum of these is equal to the overall share of people from that quartile who are married. If the model is able to match each of these moments individually, it will by definition have matched the overall marriage rate as well.

[19]We exclude the IV effect of exposure to own-race on within race marriage. There is little sex ratio variation within race groups because of high-levels of racial segregation. We show in Figure A.8 that there is approximately 36 times more sex ratio variation across versus within race groups. This is not an issue for estimating IV effects within class—there is three times more own-class versus own-race variation.

with JAX, a tool for high-performance numerical differentiation on graphical processing units (Bradbury et al., 2018). These tools can be used in other high-dimensional matching settings and we plan to post them online.

Second, we estimate a version of the model on a subsample of 150 census tracts. This brings the dimension of μ to approximately 23 million, which is computationally feasible to solve. We construct this universe of 150 census tracts to be nationally representative because our data moments are estimated on the national sample. The sampling procedure is described in detail in Appendix C.

Once μ is known, we can compute $\hat{\psi}_s(\theta)$, the model moments evaluated at θ. We repeat this process using a numerical solver, adjusting θ at each step, until the objective function in equation 10 is minimized. We defer details on the construction of the weighting matrix W, the sample distribution of types across census tracts, and other estimation protocols to Appendix D.

V.C The Impact of Reducing Segregation on Marriage Outcomes

We use our model estimates, $\hat{\theta}$, to study the impact of policies that reduce segregation in rates of interclass and interracial marriage. In addition to the estimated model parameters, we need to choose a topology of type shares across space—i.e., where people of each race and class group live—to simulate the effect of policies that reduce segregation. To do so, we use the topology of Boston to assign neighborhood race and class shares.

Our sample of chosen census tracts is shown in Figure XI Panel *A*. In Panels *B* and *C*, we plot the fraction of neighborhood residents who come from low- and high-income families. This sample clearly exhibits class based segregation. Residents from low-income families tend to be concentrated in the southeastern part of the city whereas residents from high-income families tend to live in the northern and western parts of the city.

We use our model estimates to simulate the type of policy experiment shown in Figure X and described in Section IV.C. We move one resident from a low-income family of each sex from the blue neighborhood in Figure XI Panel *D* to the red neighborhood, and one resident from a high-income family of each sex from the red neighborhood to the blue neighborhood.[20] This type of simulation is closely related to actual policies that reduce segregation by helping a small number of a city's overall residents move to higher-income neighborhoods, such as expanding access to housing vouchers. The impacts of these policies have been studied in the context of the Moving to Opportunity study and the Gautreaux Project (e.g., Bergman et al., 2019; Chyn, Collinson and Sandler, 2023; Ludwig et al., 2013).

[20] We move a racially representative resident of each class based on the national share of race groups in bottom and top quartile families. This is feasible because the matches and populations in the model are continuous rather than discrete.

To illustrate the mechanics of the model, we start by highlighting which of the model parameters in θ determine the impact of the simulation in Figure XI Panel *D* on interclass marriage and how those parameters relate to the partial equilibrium data moments. There are two key elements of $\gamma_{j,k}$. shown in equation 9.

The first is the $f(.)$ function, which determines the distance cost. The larger the distance cost, the larger the impact that this simulation will have on interclass marriage. This is because as the distance cost increases, marriage markets become more local and therefore the moves provide a larger demographic shock to the market. At the extreme of no distance cost, who lives where is irrelevant for marriage outcomes and the move has no impact on rates of interclass marriage. The fraction of the total effect on interclass marriage that is due to spillovers—i.e., impacts outside of the two treated neighborhoods such as in the Figure X example—is also closely related to the distance cost. As the distance cost increases, we approach the case where each neighborhood is an isolated marriage market and spillovers outside of the treated neighborhood, or general equilibrium effects, become small.

The second key parameter is the level of "class bias" implied by the $\chi_{j,k}$ parameters. The impact of the move on interclass marriage goes down as the preference to marry within class increases. In general, the level of bias should not have a large impact on the share of the total effect that is due to spillovers.

In Figure A.10 we show how the impact of the simulation on interclass marriage depends on distance costs. To do so, we start with a simple parameterization of $\gamma_{j,k}$ that is only a function of distance cost and therefore has no class bias. We compare two different levels of distance cost in Panels *A* and *C* by plotting the fraction of residents in each census tract who are married to a spouse from a tract in Back Bay, the northern part of the city. Marriages are spatially concentrated in Panel *A* with high distance costs and substantially more diffuse with lower distance costs in Panel *C*.

In Figure A.10 Panels *B* and *D*, we show the impact of the policy simulation on rates of interclass marriage, by neighborhood, under different parameterizations of the distance cost. Consistent with the intuition laid out above, the impact of the policy simulation on interclass marriage increases with the distance costs. The share of the total effect that is due to spillovers decreases with distance costs. We summarize these results in Figure A.11, and include two additional simulations where we introduce a substantial amount of class bias. The overall impact of the policy simulation clearly declines as class bias increases.

Having established that the impact of the policy simulation depends critically on distance costs and class bias, we next turn to understanding how these model parameters relate to our partial equilibrium data moments. Consider two moments in the context of interclass marriage between individuals from the bottom and top parent income quartiles: the effect of neighborhood exposure to one's own class group on the probability of having a within-class marriage and the effect of neighborhood exposure to the other class group on the probability of having an interclass marriage.

As distance costs increase, the overall level of these exposure effects should also increase because neighbors become a more important component of the marriage market. As class bias increases, the difference in the exposure effect for own vs. other class will widen. In Figure A.12, we show how the simulated exposure effects respond to changes in distance cost and class bias around the estimated value of $\hat{\theta}$ for individuals from high-income families. We see, reassuringly, that there are clear monotonic relationships between the model parameters and simulated data moments. Figure A.12 walks through the logic of using one moment, the own-class exposure effect, to pin down the distance cost parameter, and then using the second moment, the other-class exposure effect, to pin down the class bias.

We now turn to the impact of the policy simulation on interclass marriage at our estimated value of $\hat{\theta}$. We moved four total people in our policy simulation (described in Figure XI Panel *D*) and this move generated 0.46 new interclass marriages with 19% of the total effect coming from spillovers outside of the treated neighborhoods. The impact of the move on neighborhood-level rates of interclass marriage is shown in Figure XII. We can use the change in interclass marriage and change in exposure to compute the model implied relationship between interclass marriage and exposure. We find a gradient of 0.17, which suggests that increasing the fraction of neighbors who come from the other class group by 10pp would increase the probability of being in an interclass marriage by 1.7pp.

In Figure XIII Panel *A*, we compare the model estimated gradient of the relationship between interclass marriage and cross-class exposure to the observed relationship across commuting zones in the US. There is no ex-ante reason to expect these to be similar—the across commuting zone gradient was not included in the moments used to estimate the model, it could be contaminated by selection bias, and linearly extrapolating the model effects of small reductions in segregation could fail to capture the impact of the large differences in segregation across commuting zones. However, we find that the model estimated gradient of 0.17 closely matches the across commuting zone gradient of 0.18, implying that nearly all of the across commuting zone relationship between interclass marriage and segregation is due to the causal effect of segregation.

We repeat this policy simulation by race, moving one Black individual of each sex from the blue neighborhood in Figure XI Panel *D* to the red neighborhood, and one white individual of each sex from the red to blue neighborhood. In addition, we compare our estimates to the experimental impacts of a program that generated a similar type of desegregation as our simulation. The Gautreaux Project was a housing-desegregation program that chose Black families in Chicago, at random, to receive housing vouchers to move to low poverty neighborhoods with predominantly white residents. Chyn, Collinson and Sandler (2023), which we refer to as CCS, estimates the impact that these moves had on neighborhood racial composition and rates of interracial marriage. We scale their estimates to form an interracial marriage-exposure gradient, which we show in Figure XIII Panel *B*. The CCS estimate of 0.17 implies that a 10pp increase in the fraction of neighbors who are from the other race group increases rates of interracial marriage by 1.7pp.

The model implied interracial marriage exposure gradient is 0.12, similar to the estimate from CCS, but substantially steeper than the across commuting zone gradient in Figure XIII Panel *B*. At first glance, it might be surprising that the model produces large effects of these moves on interracial marriage despite the near zero IV effect of exposure to opposite-sex members of other race groups. In a population with heterogeneous preferences, such as the ones we model in equations 3 and 4, there will be at least some people who prefer an interracial marriage even in the presence of strong systematic ($\alpha_{j,k}$) preferences to marry within race. Distance costs combined with racial segregation will prevent these marriages from forming among the few people who desire them.

If a Black person moves to a neighborhood that had near zero Black residents previously, an interracial marriage is likely to form because segregation had been preventing the people who prefer an interracial marriage from finding a partner. This is what we find in our policy simulation, shown in Figure XIII Panel *B* and what CCS find for participants of the Gautreaux Project. An important question is whether these impacts on interracial marriage would persist if we were to undertake a larger experiment in reducing residential segregation.[21] That critically depends on the stock of people who prefer an interracial marriage, which is a function of the systematic portion of preferences, $\alpha_{j,k}$. If there are only a few people who are open to marrying across race lines, then the impact of reducing segregation will decline with each subsequent move.

V.D Total Effect of Segregation on Interclass and Interracial Marriage

We compute the rates of interclass and interracial marriage that would occur without segregation by removing the distance cost portion of preferences, the $f(.)$ term in equation 9. We add race- and class-specific intercepts to the model so that the overall group-specific marriage rates remain unchanged.

We began in Section III by showing that overall rates of interracial marriage are low. At age 30, 38.2% of white people are married to a white spouse and 2.1% of Black people are married to a white spouse, a gap of 36.1pp. In Figure XIV Panel *B* we remove distance costs, thereby removing any effect of segregation. The initial gap of 36.1pp declines to to 34pp, a reduction of only 6%.

[21] The model suggests that if we were to construct our IV estimates in neighborhoods that are entirely white, we would see a steeper relationship between market tightness and interracial marriage. We do not have power to do this in our setting with the sex ratio instrument, but our results reassuringly line up with experimental evidence from CCS.

Despite the large effects of small reductions in racial segregation, both in our model estimates and in the experimental effects of the Gautreaux Project, the overall impact of segregation on rates of interracial marriage is small.

What happens to rates of interclass marriage in a world without segregation? We show observed rates of interclass marriage and rates after removing distance costs in Figure XIV Panel *A*. At age 30, 19.0% of people from high-income families and 3.1% of people form low-income families have a spouse from a high-income family, a gap of 15.9pp. Removing distance costs reduces this gap by 41% to 9.4pp. This suggests that residential segregation alone can account for nearly half of marital homophily by class.

VI Conclusion

Marital homophily, the tendency to marry within group, is an important driver of economic inequality between race and class groups (Binder et al., 2022; Eika, Mogstad and Zafar, 2019; Mare, 1991, 2000). This type of homophily can emerge from a desire or social pressure to marry within group, a lack of exposure to individuals from different backgrounds, or a combination of both. Our results suggest low rates of interclass and interracial marriage are driven by different factors. Residential segregation, a single type of exposure, accounts for nearly half of marital homophily by class, but less than 10% of marital homophily by race.

Residential segregation has important consequences for economic inequality (e.g., Ananat, 2011; Andrews et al., 2017; Cutler and Glaeser, 1997; Massey and Denton, 1993; Wilson, 2012). We highlight an additional downstream effect of segregation: low rates of interclass marriage (Fernández and Rogerson, 2001; Kremer, 1997). A promising next step would be to study more granular data on interpersonal interactions to understand whether increasing other types of exposure, such as within-neighborhood interactions (Athey et al., 2021; Chetty et al., 2022; Moro et al., 2021), help generate additional interclass marriages.

We find that reducing residential segregation will not increase rates of interracial marriage on its own. This could be due to additional within-neighborhood segregation in interactions across race lines or strong preferences to marry within race. Estimating the relative importance of these explanations is a natural direction for future research and would improve our understanding of the factors underlying trends in interracial marriage (Bloome and Ang, 2020; Fryer, 2007).

In our model, we estimate preferences for different types of marriages and hold those preferences fixed when simulating desegregation policies. However, these preferences need not be exogenous; they may be shaped by experiences and social norms (e.g., Corno, La Ferrara and Burns, 2022). For example, living in a more racially diverse community may make a person more open to an interracial marriage. Understanding how these preferences are formed is an important next step for future research. Efforts to decrease segregation today can have downstream effects on the development of preferences for intergroup marriages, not only for adults, but also for the next generation of children growing up in these communities.

References

Abramitzky, Ran, Adeline Delavande, and Luis Vasconcelos. 2011. "Marrying up: the role of sex ratio in assortative matching." *American Economic Journal: Applied Economics*, 3(3): 124–157.

Allport, Gordon Willard, Kenneth Clark, and Thomas Pettigrew. 1954. "The nature of prejudice."

Althoff, Lukas, Harriet Brookes Gray, and Hugo Reichardt. 2022. "Intergenerational Mobility and Assortative Mating."

Ananat, Elizabeth Oltmans. 2011. "The wrong side (s) of the tracks: The causal effects of racial segregation on urban poverty and inequality." *American Economic Journal: Applied Economics*, 3(2): 34–66.

Anderson, Donald G. 1965. "Iterative procedures for nonlinear integral equations." *Journal of the ACM (JACM)*, 12(4): 547–560.

Andrews, Rodney, Marcus Casey, Bradley L Hardy, and Trevon D Logan. 2017. "Location matters: Historical racial segregation and intergenerational mobility." *Economics Letters*, 158: 67–72.

Angrist, Josh. 2002. "How do sex ratios affect marriage and labor markets? Evidence from America's second generation." *The Quarterly Journal of Economics*, 117(3): 997–1038.

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

Barlow, Fiona Kate, Stefania Paolini, Anne Pedersen, Matthew J Hornsey, Helena RM Radke, Jake Harwood, Mark Rubin, and Chris G Sibley. 2012. "The contact caveat: Negative contact predicts increased prejudice more than positive contact predicts reduced prejudice." *Personality and social Psychology bulletin*, 38(12): 1629–1643.

Bayer, Patrick, and Kerwin Kofi Charles. 2018. "Divergent paths: A new perspective on earnings differences between black and white men since 1940." *The Quarterly Journal of Economics*, 133(3): 1459–1501.

Bazzi, Samuel, Abel Brodeur, Martin Fiszbein, and Joanne Haddad. 2023. "Frontier History and Gender Norms in the United States." National Bureau of Economic Research Working Paper 31079.

Bergman, Peter, Raj Chetty, Stefanie DeLuca, Nathaniel Hendren, Lawrence F Katz, and Christopher Palmer. 2019. "Creating moves to opportunity: Experimental evidence on barriers to neighborhood choice." National Bureau of Economic Research.

Binder, Ariel J, Caroline Walker, Jonathan Eggleston, and Marta Murray-Close. 2022. "Race, Class, and Mobility in US Marriage Markets."

Bisin, Alberto, and Thierry Verdier. 2001. "The economics of cultural transmission and the dynamics of preferences." *Journal of Economic theory*, 97(2): 298–319.

Bloome, Deirdre, and Shannon Ang. 2020. "Marriage and union formation in the United States: Recent trends across racial groups and economic backgrounds." *Demography*, 57(5): 1753–1786.

Bobo, Lawrence. 2001. "Racial attitudes and relations at the close of the twentieth century." *America becoming: Racial trends and their consequences*, 1: 264–301.

Bradbury, James, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. 2018. "JAX: composable transformations of Python+NumPy programs."

Brainerd, Elizabeth. 2007. *Uncounted costs of World War II: The effects of changing sex ratios on marriage and fertility of Russian women.* National Council for Eurasian and East European Research Washington, DC.

Brummet, Quentin, et al. 2014. "Comparison of survey, federal, and commercial address data quality." *Center for Administrative Records Research and Applications Working Paper (2014-06).*

Buckles, Kasey, Joseph Price, Zachary Ward, and Haley EB Wilbert. 2023. "Family Trees and Falling Apples: Historical Intergenerational Mobility Estimates for Women and Men." National Bureau of Economic Research.

Calvo, Paula A, Ilse Lindenlaub, and Ana Reynoso. 2021. "Marriage market and labor market sorting." National Bureau of Economic Research.

Cattaneo, Matias D, Richard K Crump, Max H Farrell, and Yingjie Feng. 2019. "Binscatter regressions." *arXiv preprint arXiv:1902.09615.*

Chadwick, Laura, and Gary Solon. 2002. "Intergenerational income mobility among daughters." *American Economic Review*, 92(1): 335–344.

Charles, Kerwin Kofi, and Ming Ching Luoh. 2010. "Male incarceration, the marriage market, and female outcomes." *The Review of Economics and Statistics*, 92(3): 614–627.

Charles, Kerwin Kofi, Erik Hurst, and Alexandra Killewald. 2013. "Marital sorting and parental wealth." *Demography*, 50(1): 51–70.

Cherlin, Andrew J. 2010. *The marriage-go-round: The state of marriage and the family in America today.* Vintage.

Chetty, Raj, John N Friedman, Nathaniel Hendren, Maggie R Jones, and Sonya R Porter. 2018. "The opportunity atlas: Mapping the childhood roots of social mobility." National Bureau of Economic Research.

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

Chetty, Raj, Nathaniel Hendren, Maggie R Jones, and Sonya R Porter. 2020. "Race and economic opportunity in the United States: An intergenerational perspective." *The Quarterly Journal of Economics*, 135(2): 711–783.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. "Where is the land of opportunity? The geography of intergenerational mobility in the United States." *The Quarterly Journal of Economics*, 129(4): 1553–1623.

Chetty, Raj, Will Dobbie, Benjamin Goldman, Sonya Porter, and Cristal Yang. 2023. "Growing Class Gaps, Shrinking Race Gaps: Economic and Sociological Mechanisms Underlying Recent Trends in Intergenerational Mobility."

Chiappori, Pierre-André, Monica Costa Dias, and Costas Meghir. 2018. "The marriage market, labor supply, and education choice." *Journal of Political Economy*, 126(S1): S26–S72.

Chiappori, Pierre-Andre, Murat Iyigun, and Yoram Weiss. 2009. "Investment in schooling and the marriage market." *American Economic Review*, 99(5): 1689–1713.

Chiappori, Pierre-André, Sonia Oreffice, and Climent Quintana-Domeque. 2012. "Fatter attraction: anthropometric and socioeconomic matching on the marriage market." *Journal of Political Economy*, 120(4): 659–695.

Chiappori, Pierre-André, Sonia Oreffice, and Climent Quintana-Domeque. 2018. "Bidimensional matching with heterogeneous preferences: education and smoking in the marriage market." *Journal of the European Economic Association*, 16(1): 161–198.

Choo, Eugene, and Aloysius Siow. 2006. "Who marries whom and why." *Journal of political Economy*, 114(1): 175–201.

Chyn, Eric, Robert Collinson, and Danielle Sandler. 2023. "The Long-Run Effects of Residential Racial Desegregation Programs: Evidence from Gautreaux."." Working Paper.

Ciscato, Edoardo, and Simon Weber. 2020. "The role of evolving marital preferences in growing income inequality." *Journal of Population Economics*, 33(1): 307–347.

Clark, Gregory, and Neil Cummins. 2022. "Assortative Mating and the Industrial Revolution: England, 1754-2021."

Corno, Lucia, Eliana La Ferrara, and Justine Burns. 2022. "Interaction, stereotypes, and performance: Evidence from South Africa." *American Economic Review*, 112(12): 3848–3875.

Cox, Oliver C. 1940. "Sex ratio and marital status among Negroes." *American sociological review*, 5(6): 937–947.

Cullen, Julie Berry, and Jonathan Gruber. 2000. "Does unemployment insurance crowd out spousal labor supply?" *Journal of labor Economics*, 18(3): 546–572.

Cutler, David M, and Edward L Glaeser. 1997. "Are ghettos good or bad?" *The Quarterly Journal of Economics*, 112(3): 827–872.

Easterlin, Richard A. 1961. "The American Baby Boom in Historical Perspective." *The American Economic Review*, 51(5): 869–911.

Eika, Lasse, Magne Mogstad, and Basit Zafar. 2019. "Educational assortative mating and household income inequality." *Journal of Political Economy*, 127(6): 2795–2835.

Eriksson, Katherine, Gregory T Niemesh, Myera Rashid, and Jacqueline Craig. 2023. "Marriage and the Intergenerational Mobility of Women: Evidence from Marriage Certificates 1850-1920."

Ermisch, John, Marco Francesconi, and Thomas Siedler. 2006. "Intergenerational mobility and marital sorting." *The Economic Journal*, 116(513): 659–679.

Espín-Sánchez, José-Antonio, Joseph P Ferrie, and Christopher Vickers. 2023. "Women and the Econometrics of Family Trees." National Bureau of Economic Research.

Fernández, Raquel, and Richard Rogerson. 2001. "Sorting and long-run inequality." *The Quarterly Journal of Economics*, 116(4): 1305–1341.

Fisman, Raymond, Sheena S Iyengar, Emir Kamenica, and Itamar Simonson. 2008. "Racial preferences in dating." *The Review of Economic Studies*, 75(1): 117–132.

Foster, Thomas B. 2017. "Decomposing American immobility: Compositional and rate components of interstate, intrastate, and intracounty migration and mobility decline." *Demographic Research*, 37: 1515–1548.

Fox, Jeremy T. 2010. "Identification in matching games." *Quantitative Economics*, 1(2): 203–254.

Fryer, Roland G. 2007. "Guess who's been coming to dinner? Trends in interracial marriage over the 20th century." *Journal of Economic Perspectives*, 21(2): 71–90.

Galichon, Alfred, and Bernard Salanié. 2022. "Cupid's invisible hand: Social surplus and identification in matching models." *The Review of Economic Studies*, 89(5): 2600–2629.

Glaeser, Edward L, and Joseph Gyourko. 2002. "The impact of zoning on housing affordability."

Goni, Marc. 2022. "Assortative Matching at the Top of the Distribution: Evidence from the World's Most Exclusive Marriage Market." *American Economic Journal: Applied Economics*, 14(3): 445–487.

Gordon, Nora, and Sarah Reber. 2018. "The effects of school desegregation on mixed-race births." *Journal of Population Economics*, 31: 561–596.

Graham, Bryan S. 2013. "Comparative static and computational methods for an empirical one-to-one transferable utility matching model." In *Structural Econometric Models*. Vol. 31, 153–181. Emerald Group Publishing Limited.

Greenwood, Jeremy, Nezih Guner, Georgi Kocharkov, and Cezar Santos. 2014. "Marry your like: Assortative mating and income inequality." *American Economic Review*, 104(5): 348–53.

Groves, Ernest Rutherford, and William Fielding Ogburn. 1928. *American marriage and family relationships.* Ayer Publishing.

Gualdani, Cristina, and Shruti Sinha. 2020. "Partial identification in matching models for the marriage market."

Hall, Matthew, John Iceland, and Youngmin Yi. 2019. "Racial separation at home and work: Segregation in residential and workplace settings." *Population Research and Policy Review*, 38: 671–694.

Haltiwanger, John C, Mark J Kutzbach, Giordano E Palloni, Henry Pollakowski, Matthew Staiger, and Daniel Weinberg. 2020. "The children of HOPE VI demolitions: National evidence on labor market outcomes." National Bureau of Economic Research.

Hitsch, Günter J, Al Hortaçsu, and Dan Ariely. 2010. "Matching and sorting in online dating." *American Economic Review*, 100(1): 130–163.

Hoxby, Caroline M. 2000. "Peer effects in the classroom: Learning from gender and race variation."

Keller, Wolfgang, and Carol H Shiue. 2023. "Intergenerational Mobility of Daughters and Marital Sorting: New Evidence from Imperial China." National Bureau of Economic Research Working Paper 31695.

Kennedy, Sheela, and Larry Bumpass. 2008. "Cohabitation and children's living arrangements: New estimates from the United States." *Demographic research*, 19: 1663.

Kirkham, Elyssa. 2019. "2019 survey on dating and distance: how far are people willing to look for love?" *Ladders*.

Kreider, Rose M, Lydia Anderson, Thomas Gryn, and Chanell Washington. 2023. "Same-Sex Couples Made Up 1.7% of All Coupled Households." Accessed on June 28, 2023.

Kremer, Michael. 1997. "How much does sorting increase inequality?" *The Quarterly Journal of Economics*, 112(1): 115–139.

Lafortune, Jeanne. 2013. "Making yourself attractive: Pre-marital investments and the returns to education in the marriage market." *American Economic Journal: Applied Economics*, 5(2): 151–178.

Liebler, Carolyn A, Sonya R Porter, Leticia E Fernandez, James M Noon, and Sharon R Ennis. 2017. "America's churning races: Race and ethnicity response changes between census 2000 and the 2010 census." *Demography*, 54(1): 259–284.

Logan, John R, Andrew Foster, Hongwei Xu, and Wenquan Zhang. 2020. "Income segregation: Up or down, and for whom?" *Demography*, 57(5): 1951–1974.

Low, Corinne. 2021. "The Human Capital-Reproductive Capital Tradeoff in Marriage Market Matching." Mimeo.

Ludwig, Jens, Greg J Duncan, Lisa A Gennetian, Lawrence F Katz, Ronald C Kessler, Jeffrey R Kling, and Lisa Sanbonmatsu. 2013. "Long-term neighborhood effects on low-income families: Evidence from Moving to Opportunity." *American economic review*, 103(3): 226–231.

Lundberg, Shelly. 1985. "The added worker effect." *Journal of Labor Economics*, 3(1, Part 1): 11–37.

Lundberg, Shelly, Robert A Pollak, and Jenna Stearns. 2016. "Family inequality: Diverging patterns in marriage, cohabitation, and childbearing." *Journal of Economic perspectives*, 30(2): 79–102.

Manning, Wendy D, and Jessica A Cohen. 2012. "Premarital cohabitation and marital dissolution: An examination of recent marriages." *Journal of Marriage and Family*, 74(2): 377–387.

Mare, Robert. 2000. "Assortative mating, intergenerational mobility, and educational inequality." *UCLA CCPR Population Working Papers*.

Mare, Robert D. 1991. "Five decades of educational assortative mating." *American sociological review*, 15–32.

Massey, Douglas S, and Nancy A Denton. 1988. "The dimensions of residential segregation." *Social forces*, 67(2): 281–315.

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

Mazumder, Bhashkar. 2005. "Fortunate sons: New estimates of intergenerational mobility in the United States using social security earnings data." *Review of Economics and Statistics*, 87(2): 235–255.

McFadden, Daniel. 1989. "A method of simulated moments for estimation of discrete response models without numerical integration." *Econometrica: Journal of the Econometric Society*, 995–1026.

McPherson, Miller, Lynn Smith-Lovin, and James M Cook. 2001. "Birds of a feather: Homophily in social networks." *Annual review of sociology*, 415–444.

Merlino, Luca Paolo, Max Friedrich Steinhardt, and Liam Wren-Lewis. 2019. "More than just friends? School peers and adult interracial relationships." *Journal of Labor Economics*, 37(3): 663–713.

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

Mourifié, Ismael, and Aloysius Siow. 2021. "The Cobb-Douglas marriage matching function: marriage matching with peer and scale effects." *Journal of Labor Economics*, 39(S1): S239–S274.

Olivetti, Claudia, M Daniele Paserman, Laura Salisbury, and E Anna Weber. 2020. "Who married,(to) whom, and where? Trends in marriage in the United States, 1850-1940." National Bureau of Economic Research.

Paolini, Stefania, Jake Harwood, and Mark Rubin. 2010. "Negative intergroup contact makes group memberships salient: Explaining why intergroup conflict endures." *Personality and Social Psychology Bulletin*, 36(12): 1723–1738.

Pencavel, John. 1998. "Assortative mating by schooling and the work behavior of wives and husbands." *The American Economic Review*, 88(2): 326–329.

Popkin, Susan J. 2004. "A decade of HOPE VI: Research findings and policy challenges."

Reardon, Sean F, and Kendra Bischoff. 2011. "Income inequality and income segregation." *American journal of sociology*, 116(4): 1092–1153.

Rosenfeld, Michael J. 2017. "Marriage, choice, and couplehood in the age of the internet." *Sociological Science*, 4: 490–510.

Rosenfeld, Michael J, and Reuben J Thomas. 2012. "Searching for a mate: The rise of the Internet as a social intermediary." *American Sociological Review*, 77(4): 523–547.

Rosenfeld, Michael J, Reuben J Thomas, and Sonia Hausen. 2019. "Disintermediating your friends: How online dating in the United States displaces other ways of meeting." *Proceedings of the National Academy of Sciences*, 116(36): 17753–17758.

Rothwell, Jonathan, and Douglas S Massey. 2009. "The effect of density zoning on racial segregation in US urban areas." *Urban Affairs Review*, 44(6): 779–806.

Rothwell, Jonathan T, and Douglas S Massey. 2010. "Density zoning and class segregation in US metropolitan areas." *Social science quarterly*, 91(5): 1123–1143.

Schwartz, Christine R. 2010. "Earnings inequality and the changing association between spouses' earnings." *American journal of sociology*, 115(5): 1524–1557.

Sinha, Shruti. 2018. "Identification in One-to-One Matching Models with Nonparametric Unobservables."

Siow, Aloysius. 2015. "Testing Becker's theory of positive assortative matching." *Journal of Labor Economics*, 33(2): 409–441.

Solon, Gary. 1992. "Intergenerational income mobility in the United States." *The American Economic Review*, 393–408.

South, Scott J, and Kim M Lloyd. 1992. "Marriage opportunities and family formation: Further implications of imbalanced sex ratios." *Journal of Marriage and the Family*, 440–451.

Sprung-Keyser, Ben, Nathaniel Hendren, and Sonya Porter. 2022. "The Radius of Economic Opportunity: Evidence from Migration and Local Labor Markets."

Wagner, Deborah, and Mary Lane. 2014. "The person identification validation system (PVS): applying the Center for Administrative Records Research and Applications'(CARRA) record linkage software." Center for Economic Studies, US Census Bureau.

Wilson, William Julius. 2012. *The truly disadvantaged: The inner city, the underclass, and public policy.* University of Chicago Press.

Wong, Linda Y. 2002. "Why Do Only 5.5% of Black Men Marry White Women? The Impact of the Mating Taboo, Courtship Opportunities, and Individual Endowments." *The Impact of the Mating Taboo, Courtship Opportunities, and Individual Endowments (December 2002).*

Zaremba, Krzysztof. 2022. "Effects of Intra-Couple Bargaining Power on Maternal and Neonatal Health."

TABLE I: Summary Statistics by Class

	Pooled (1)	Bottom 25% (2)	Quartile 2 (3)	Quartile 3 (4)	Top 25% (5)
A. Parent Characteristics					
Median Parent Household Income ($)	55170	15010	39270	73850	136400
Mean Parent Household Income Percentile Rank	0.500	0.125	0.375	0.625	0.875
Both Parents Present in Household	0.673	0.392	0.574	0.805	0.919
Parents are Same Race	0.918	0.899	0.909	0.920	0.929
B. Child Earnings at age 30					
Median Household Income ($)	34430	19600	29130	40390	56150
Mean Household Income Percentile Rank	0.500	0.375	0.454	0.540	0.631
Median Individual Income ($)	26090	15330	22430	29740	39980
Mean Individual Income Percentile Rank	0.500	0.393	0.460	0.529	0.618
Employed (Individual Income > 0)	0.801	0.720	0.787	0.834	0.863
C. Child Location at Age 27					
Has a Non-Missing Address	0.869	0.787	0.853	0.902	0.934
After Imputation	0.998	0.996	0.998	0.999	1.000
Lives within 10 Census Tracts of Age 18 Address	0.379	0.370	0.393	0.403	0.353
Lives within 50 Census Tracts of Age 18 Address	0.510	0.521	0.534	0.532	0.453
Lives in Age 18 County	0.560	0.594	0.584	0.561	0.501
Lives in Age 18 Commuting Zone	0.672	0.701	0.697	0.676	0.618
D. Child Marriage Outcomes at age 30					
Ever Married	0.451	0.365	0.434	0.502	0.500
Currently Married	0.350	0.239	0.317	0.405	0.437
Mean Spouse Parent Income Rank	0.580	0.454	0.514	0.582	0.680
Mean Spouse Individual Income Rank (Age 30)	0.603	0.523	0.559	0.601	0.671
Has a Spouse With:					
Non-missing Parent Income	0.297	0.180	0.261	0.355	0.392
Same Parent Income Quartile	0.107	0.050	0.070	0.117	0.190
Fully Random Benchmark	0.074	0.074	0.074	0.074	0.074
Random Spouse Benchmark	0.074	0.035	0.061	0.100	0.116
Number of Children (1,000s)	31060	7765	7765	7765	7765
Fraction in each Parent Income Group		0.250	0.250	0.250	0.250

Notes: This table presents summary statistics on the sample and marriage outcomes by class (parent income). See Table II for analogous statistics broken down by race. Parent income ranks are defined relative to all parents with children in the same birth cohort. All statistics are based on the primary analysis sample (individuals in the 1982-1989 birth cohorts). Panels A and B present descriptive statistics on parents' and individual's incomes, respectively. Panel C provides information on the availability of address data and the moving patterns of individuals in adulthood. Panel D presents descriptives on marriage outcomes measured at age 30. All monetary values are measured in 2015 dollars. Parent household income is defined as the mean when the child is 13-17, assigning zeros for non-filers. Mean spouse parent income rank is calculated only for those who are married to a spouse with non-missing parent income. See Section II for more detail on the matching procedure. Spouse individual income is calculated in the year the individual is 30, as opposed to when the spouse is 30, if the two are born in different years, and is calculated only for individuals who are married. The fully random benchmark is the overall marriage rate multiplied by the fraction of the population from each class group. The random spouse benchmark is equal to the actual marriage rate for each class group multiplied by the fraction of the married population that comes from each class group. For both benchmarks, we include those with missing class the population, which we estimate as the share of spouses with missing parent income data (15%). All values in this and all subsequent tables and figures have been rounded as part of the Census disclosure avoidance protocol. All statistics in this and subsequent tables and figures cleared under Census DRB release authorizations CBDRB-FY23-CES014-009, CBDRB-FY23-0492.

TABLE II: Summary Statistics by Race

	Pooled (1)	White (2)	Black (3)	Asian (4)	Hispanic (5)	AIAN (6)	Other (7)
A. Parent Characteristics							
Median Parent Household Income ($)	55170	72280	29610	56480	34610	33400	52120
Mean Parent Household Income Percentile Rank	0.500	0.581	0.332	0.509	0.375	0.362	0.485
Both Parents Present in Household	0.673	0.799	0.291	0.806	0.545	0.554	0.647
Parents are Same Race	0.918	0.960	0.908	0.901	0.775	0.582	0.475
B. Child Earnings at age 30							
Median Household Income ($)	34430	42650	18470	46640	29900	19270	31900
Mean Household Income Percentile Rank	0.500	0.554	0.356	0.578	0.458	0.372	0.480
Median Individual Income ($)	26090	30080	17230	37510	24160	14340	25170
Mean Individual Income Percentile Rank	0.500	0.535	0.409	0.596	0.475	0.386	0.492
Employed (Individual Income > 0)	0.801	0.822	0.791	0.830	0.798	0.720	0.803
C. Child Location at Age 27							
Has a Non-Missing Address	0.869	0.903	0.823	0.884	0.860	0.745	0.856
After Imputation	0.998	0.999	0.999	0.999	0.998	0.987	0.999
Lives within 10 Census Tracts of Age 18 Address	0.379	0.377	0.349	0.400	0.399	0.411	0.345
Lives within 50 Census Tracts of Age 18 Address	0.510	0.508	0.492	0.491	0.531	0.553	0.466
Lives in Age 18 County	0.560	0.519	0.612	0.603	0.662	0.539	0.548
Lives in Age 18 Commuting Zone	0.672	0.640	0.721	0.705	0.754	0.635	0.650
D. Child Marriage Outcomes at age 30							
Ever Married	0.451	0.533	0.197	0.378	0.423	0.393	0.402
Currently Married	0.350	0.433	0.117	0.316	0.289	0.256	0.299
Mean Spouse Parent Income Rank	0.580	0.607	0.428	0.572	0.463	0.458	0.555
Mean Spouse Individual Income Rank (Age 30)	0.603	0.612	0.534	0.686	0.566	0.521	0.589
Has a Spouse With:							
Non-missing Race	0.331	0.421	0.107	0.264	0.246	0.244	0.278
Same Race	0.274	0.382	0.072	0.148	0.144	0.072	0.043
Fully Random Benchmark	0.154	0.218	0.048	0.012	0.050	0.003	0.009
Random Spouse Benchmark	0.210	0.319	0.005	0.010	0.033	0.002	0.006
Number of Children (1,000s)	31060	19080	4176	1071	4410	261	801
Fraction in each Race Group		0.609	0.133	0.034	0.141	0.008	0.026

Notes: This table presents summary statistics analogous to Table I, but broken down by race. All racial groups except Hispanics exclude individuals of Hispanic ethnicity. Column 6 includes individuals with two or more race groups listed. Note that the pooled row excludes includes the missing race group which is approx 5% of the sample; sample shares listed at the bottom do not sum to 1. The benchmarks are defined as in Table I. In these benchmarks, the population shares account for spouses who have missing race data (5% of spouses). See the notes to Table I for additional details.

TABLE III: Marriage Outcomes by Class Relative to Random Matching Benchmark

Parent Income Group		Fraction Married (1)	Fraction with Spouse from Parent Income Group at Age 30: Bottom 25% (2)	Quartile 2 (3)	Quartile 3 (4)	Top 25% (5)	Missing Par. Inc. (6)
Bottom 25%	Truth	0.239	0.050	0.052	0.047	0.031	0.059
	Fully Random	0.350	0.074	0.074	0.074	0.074	0.053
	Random Spouse	0.239	0.035	0.046	0.059	0.063	0.036
Quartile 2	Truth	0.317	0.052	0.070	0.079	0.059	0.056
	Fully Random	0.350	0.074	0.074	0.074	0.074	0.053
	Random Spouse	0.317	0.046	0.061	0.078	0.084	0.048
Quartile 3	Truth	0.405	0.048	0.080	0.117	0.111	0.050
	Fully Random	0.350	0.074	0.074	0.074	0.074	0.053
	Random Spouse	0.405	0.059	0.078	0.100	0.108	0.061
Top 25%	Truth	0.437	0.031	0.060	0.111	0.190	0.045
	Fully Random	0.350	0.074	0.074	0.074	0.074	0.053
	Random Spouse	0.437	0.063	0.084	0.108	0.116	0.066

Notes: This table presents observed rates of interclass marriage, measured at age 30, and two benchmarks. For each of the four class quartiles, we calculate the overall marriage rate, the fraction who have a spouse whose own parent(s) were in each of the four quartiles, as well as the fraction who have a spouse who is missing parent income. These spouses cannot be matched to parents using the procedure described in Section II. The fully random benchmark is the overall marriage rate multiplied by the fraction of the population from each class group. The random spouse benchmark is equal to the actual marriage rate for each class group multiplied by the fraction of the married population that comes from each class group. For both benchmarks, we include those with missing class the population, which we estimate as the share of spouses with missing parent income data.

TABLE IV: Marriage Outcomes by Race Relative to Random Matching Benchmark

			Fraction with Spouse from Race Group at Age 30						
Race Group		Fraction Married (1)	White (2)	Black (3)	Asian (4)	Hispanic (5)	AIAN (6)	Other (7)	Missing Race (8)
White	Truth	0.433	0.382	0.005	0.005	0.020	0.002	0.007	0.012
	Fully Random	0.358	0.218	0.048	0.012	0.050	0.003	0.009	0.018
	Random Spouse	0.433	0.319	0.019	0.013	0.049	0.003	0.009	0.021
Black	Truth	0.117	0.021	0.072	0.001	0.008	0.000	0.004	0.009
	Fully Random	0.358	0.218	0.048	0.012	0.050	0.003	0.009	0.018
	Random Spouse	0.117	0.086	0.005	0.004	0.013	0.001	0.002	0.006
Asian	Truth	0.316	0.077	0.006	0.148	0.017	0.001	0.015	0.052
	Fully Random	0.358	0.218	0.048	0.012	0.050	0.003	0.009	0.018
	Random Spouse	0.316	0.233	0.014	0.010	0.036	0.002	0.007	0.016
Hispanic	Truth	0.289	0.083	0.008	0.005	0.144	0.001	0.006	0.043
	Fully Random	0.358	0.218	0.048	0.012	0.050	0.003	0.009	0.018
	Random Spouse	0.289	0.213	0.013	0.009	0.033	0.002	0.006	0.014
AIAN	Truth	0.256	0.132	0.007	0.003	0.020	0.072	0.010	0.012
	Fully Random	0.358	0.218	0.048	0.012	0.050	0.003	0.009	0.018
	Random Spouse	0.256	0.189	0.011	0.008	0.029	0.002	0.005	0.013
Other	Truth	0.299	0.159	0.021	0.021	0.031	0.003	0.043	0.021
	Fully Random	0.358	0.218	0.048	0.012	0.050	0.003	0.009	0.018
	Random Spouse	0.299	0.220	0.013	0.009	0.034	0.002	0.006	0.015

Notes: This table presents observed rates of interracial marriage, measured at age 30, and two benchmarks. For each of the six race groups, we calculate the marriage rate, the fraction who have a spouse from each of the race groups, as well as what fraction have a spouse who is missing race. Spouses are assigned race following the procedure in Section II, regardless of whether or not they can be assigned to parents. The rates of missing spouse race are substantially lower than mising spouse class, for this reason. The fully random benchmark is the overall marriage rate multiplied by the fraction of the population from each race group. The random spouse benchmark is equal to the actual marriage rate for each race group multiplied by the fraction of the married population that comes from each race group. For both benchmarks, we include those with missing race the population, which we estimate as the share of spouses with missing race data. Note that to construct the benchmarks in this table, unlike Table II which uses the full sample, we exclude individuals who are themselves missing race (approximately 5% of individuals). They are computed only using individuals who are white, Black, Asian, Hispanic, AIAN, or in the other race group.

TABLE V: Distance from Eventual Spouse

	By Age			By Year Prior to Marriage					Excluding Couples at Same Address: By Age			Excluding Couples at Same Address: By Year Prior to Marriage				
	0	10	18	1	3	5	10	20	0	10	18	1	3	5	10	20
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
A. Percent of Couples that Ever Lived in Nearest Census Tracts																
100 Census Tracts	0.29	0.40	0.51	0.96	0.85	0.73	0.52	0.36	0.29	0.40	0.51	0.73	0.70	0.65	0.51	0.36
50 Census Tracts	0.23	0.34	0.45	0.96	0.81	0.68	0.46	0.30	0.23	0.34	0.45	0.66	0.63	0.58	0.45	0.30
25 Census Tracts	0.18	0.28	0.39	0.95	0.78	0.62	0.40	0.24	0.18	0.28	0.38	0.58	0.55	0.50	0.38	0.24
10 Census Tracts	0.12	0.20	0.29	0.93	0.73	0.54	0.30	0.17	0.12	0.20	0.29	0.46	0.44	0.39	0.29	0.17
B. Percent of Couples that Ever Lived in Radius																
25 Miles	0.38	0.47	0.58	0.97	0.88	0.78	0.60	0.44	0.37	0.47	0.58	0.80	0.78	0.73	0.59	0.44
10 Miles	0.24	0.34	0.44	0.95	0.82	0.68	0.46	0.30	0.24	0.34	0.44	0.67	0.64	0.59	0.45	0.30
5 Miles	0.14	0.23	0.32	0.94	0.75	0.57	0.33	0.19	0.14	0.22	0.31	0.52	0.49	0.44	0.32	0.19
C. Percent of Couples that Ever Lived in Same Geographic Area																
State	0.56	0.65	0.75	0.99	0.94	0.89	0.76	0.62	0.56	0.65	0.75	0.91	0.89	0.86	0.75	0.62
Commuting Zone	0.40	0.49	0.59	0.97	0.89	0.79	0.61	0.46	0.40	0.49	0.59	0.81	0.79	0.74	0.60	0.46
County	0.27	0.37	0.47	0.96	0.83	0.70	0.49	0.33	0.27	0.37	0.47	0.70	0.67	0.62	0.48	0.33
Census Tract	0.03	0.05	0.09	0.89	0.60	0.36	0.11	0.04	0.03	0.05	0.09	0.18	0.17	0.14	0.09	0.04

Notes: This table presents statistics on the geographic proximity of married couples in childhood and adulthood. All statistics are constructed from a dataset that contains one row per couple. The ages are evaluated relative to the younger member of the couple, if the individuals were not born in the same year. Parent addresses are used for ages 18 and below. Addresses at later ages are defined using the individual's own addresses. Missing years are imputed, separately for childhood and adulthood locations. For more information on the construction of the address panel, see Section II. Columns 1-3 present the fraction of eventual couples that had ever lived within a defined geographic area by age 0, 10, and 18, respectively. Columns 4-7 present analogous statistics, but evaluated at particular points in time relative to the year the couple marries. In Columns 1-7, couples may be near each other because they are living together in the same home. Columns 8-14 repeat the calculations, excluding years in which couples reside at the exact same address, defined using the MAFID.

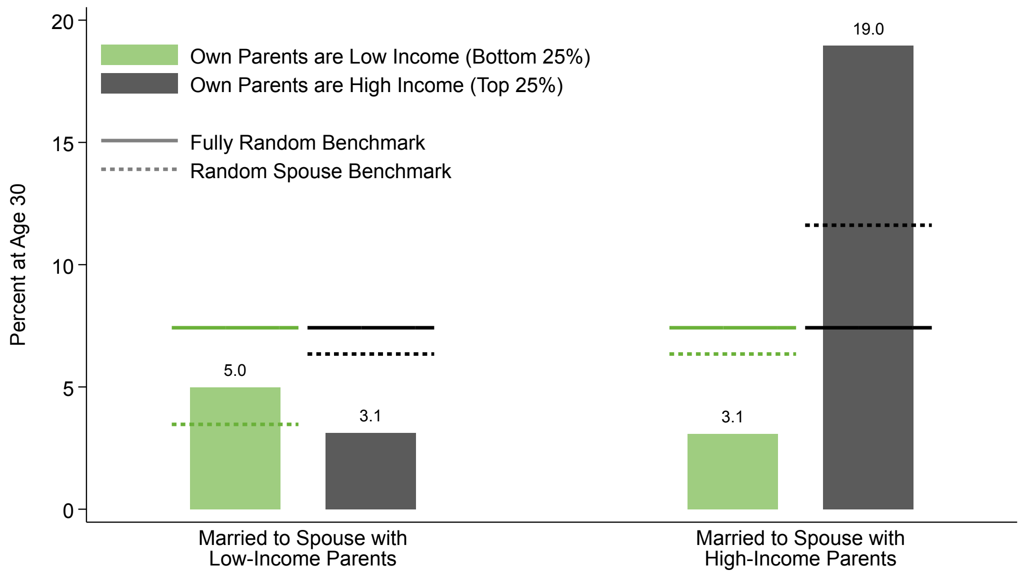
TABLE VI: Effect of Market Tightness Among Other Race or Class Group on Intergroup Marriage and Overall Marriage Rates

	First Stage	IV Estimates: Y = Married to Other Group			IV Estimates: Y = Married
	(1)	(2)	(3)	(4)	(5)
A. Estimates by Class					
Bottom & Top 25%	0.173	0.132	0.113	0.140	0.027
	(0.006)	(0.027)	(0.038)	(0.034)	(0.083)
Observations (1,000s)	15070	15070	7062	5399	15070
B. Estimates by Race					
White & Black	0.035	0.001	-0.022	-0.001	0.035
	(0.001)	(0.018)	(0.027)	(0.024)	(0.173)
Observations (1,000s)	22660	22660	11750	8987	22660
Baseline Controls	Yes	Yes	Yes	Yes	Yes
Family FEs	No	No	Yes	No	No
Family x Childhood Tract FEs	No	No	No	Yes	No

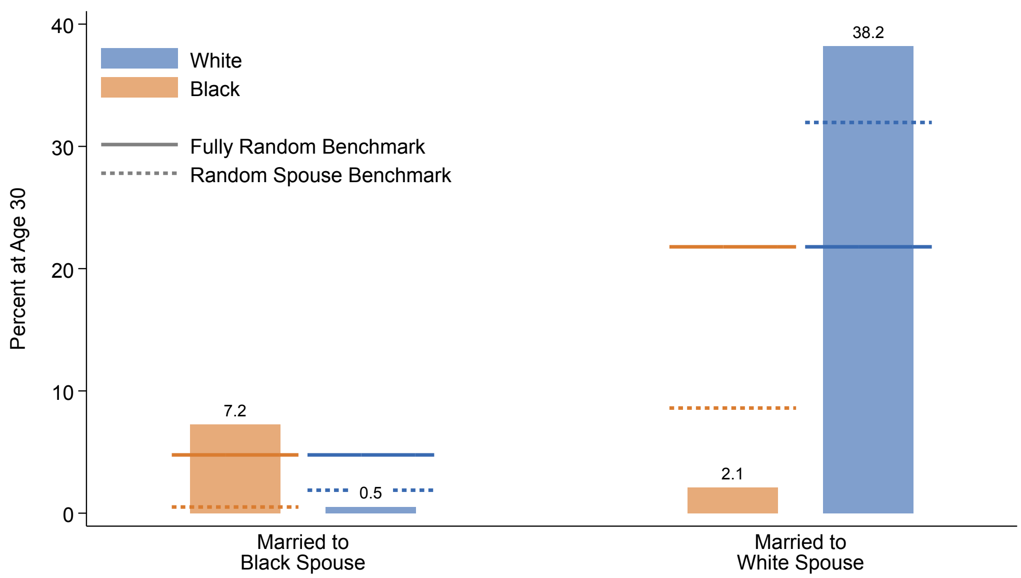
Notes: This table presents regression coefficients for the effect of exposure to different class and race groups on marriage outcomes. In Panel A, we use a sample of individuals from low-income (bottom quartile) and high-income (top quartile) families. In Panel B, we use a sample of white and Black individuals. The baseline controls consist of class (race) by sex by age 18 Census tract fixed effects and linear controls for the fraction of neighbors who are own-sex and the fraction of neighbors in the other class (race) group, interacted with class (race) and sex fixed effects. In Column (1), we present the first stage coefficients, where the endogenous treatment variable is the market tightness measure defined in Section IV.B and the instrument is the sex-ratio variation described in IV.A. In Column (2), we present the baseline IV coefficients shown in Figure VIII. In Panel A, the outcome is an indicator for having spouse from the other class group (i.e. a spouse from a low-income family if the individual is from a high-income family and vice versa). In Panel B, the outcome is an indicator for having a spouse from the other race (i.e. a white spouse if the individual is Black and vice versa). In Column (3), we restrict the sample to a set of siblings and include family fixed instead of the class (race) by sex by age 18 Census tract fixed effects. In Column (4), we further restrict to siblings who lived in the same Census tract at age 18. Finally, in Column (5) we report outcomes on the extensive margin using an indicator for being married at all. In each specification, standard errors are clustered at the level of a person's age 18 county.

FIGURE I: Marital Homophily by Class and Race

A: Rates of Interclass vs. Within-Class Marriage



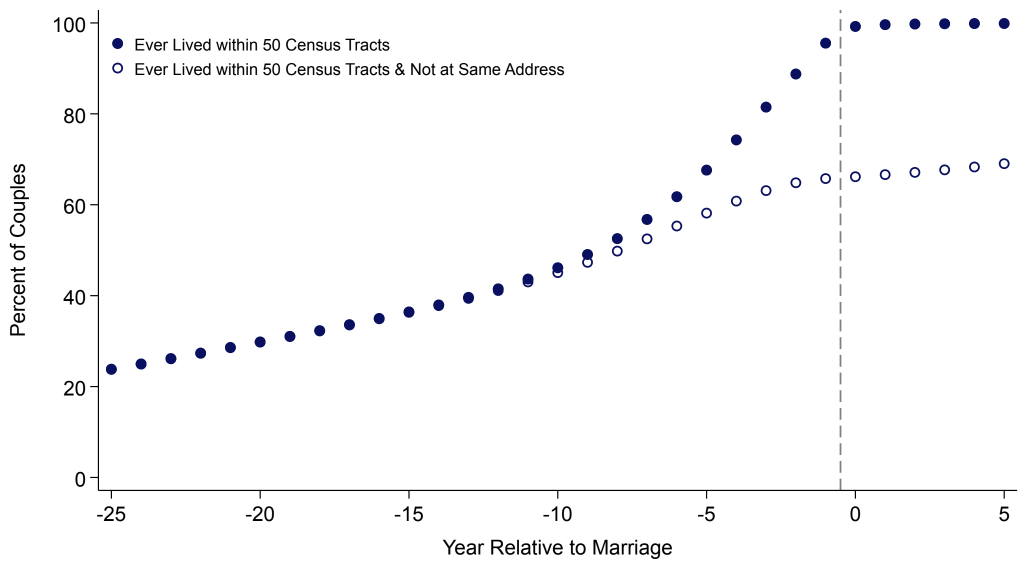
B: Rates of Interracial vs. Within-Race Marriage



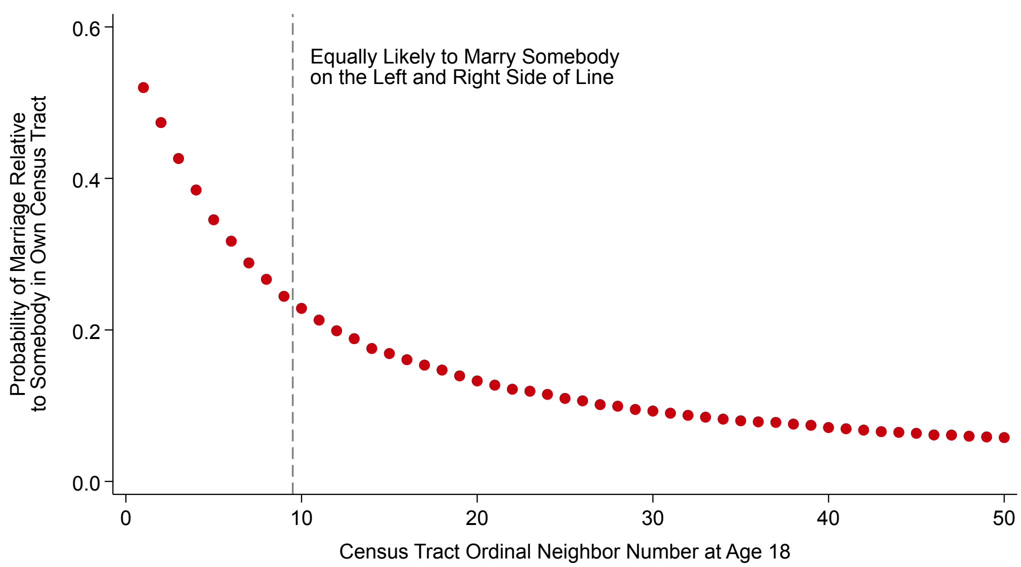
Notes: This figure summarizes key patterns of marital homophily by class (parent income) and race. In Panel *A*, we plot the fraction of individuals who are married at age 30 to a spouse from a low-income (bottom parent quartile) and high-income (top quartile) family, separately for those who themselves come from low- or high-income families. We include unmarried people in the calculation. In Panel *B*, we construct analogous statistics, using a sample of white and Black individuals. We include two benchmarks. The fully random benchmark is the overall marriage rate multiplied by the fraction of the population from each class or race group. The random spouse benchmark is equal to the actual marriage rate for each class or race group multiplied by the fraction of the married population that comes from each class or race group. For both benchmarks, we include those with missing class or race in the population, which we estimate as the share of spouses with missing data. Similar statistics for other class and race groups are available in Tables III and IV, respectively.

FIGURE II: The Role of Neighborhoods in Marriage Markets

A: Distance From Spouse by Years from Marriage



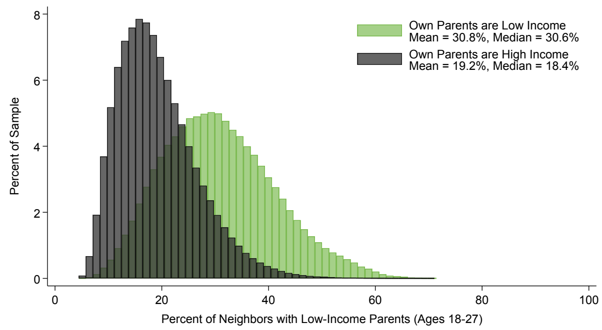
B: Marriage Probability Decay by Distance



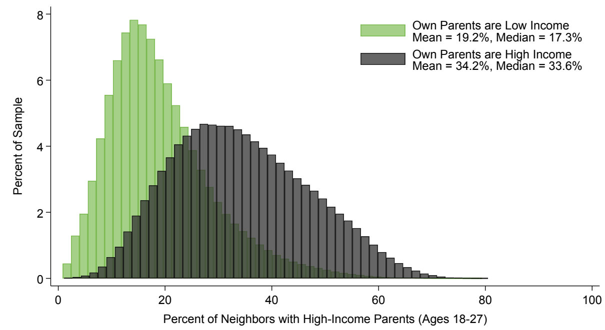
Notes: This figure presents statistics on the local nature of marriage markets. In Panel *A*, we plot the fraction of couples who had ever lived within 50 Census tracts by a particular year relative to the year they married. The solid circles includes individuals who lived near each other because they lived together at the same address. In the hollow cirlces, we plot the fraction who had ever lived within 50 tracts, excluding years in which couples reside at the exact same address, defined by MAFID. For similar statistics at fixed ages and for different geographic areas, see Table V. In Panel *B*, we plot the probability that individuals marry a person in each of the 50 nearest Census tracts to their age 18 address, relative to the probability that they marry a person in their exact age 18 Census tract. The dashed vertical line between 9 and 10 signifies that individuals are as likely to marry someone from their own tract or the nearest 9 neighbors as they are to marry somone from nearest neighbors 10-50 combined. A version of this plot with separate probabilities by tercile of population density is available in Figure A.2 Panel *A*. That version serves as the decay weights used to weight neighbors as described in Section III.C.

FIGURE III: Segregation in Adult Neighborhoods
Distribution of Exposure to Race and Class Groups

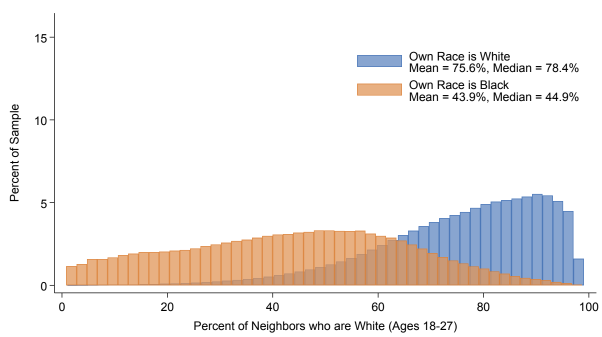
A: Low-Income (Bottom 25%) Parents



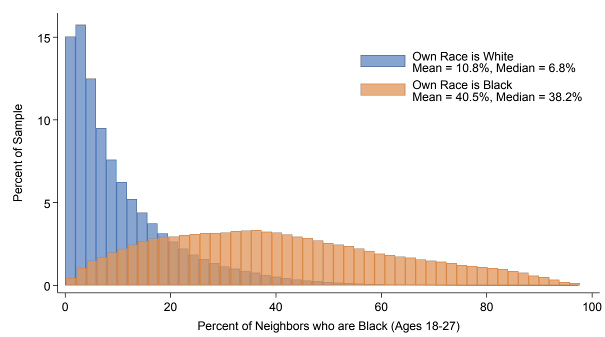
B: High-Income (Top 25%) Parents



C: White

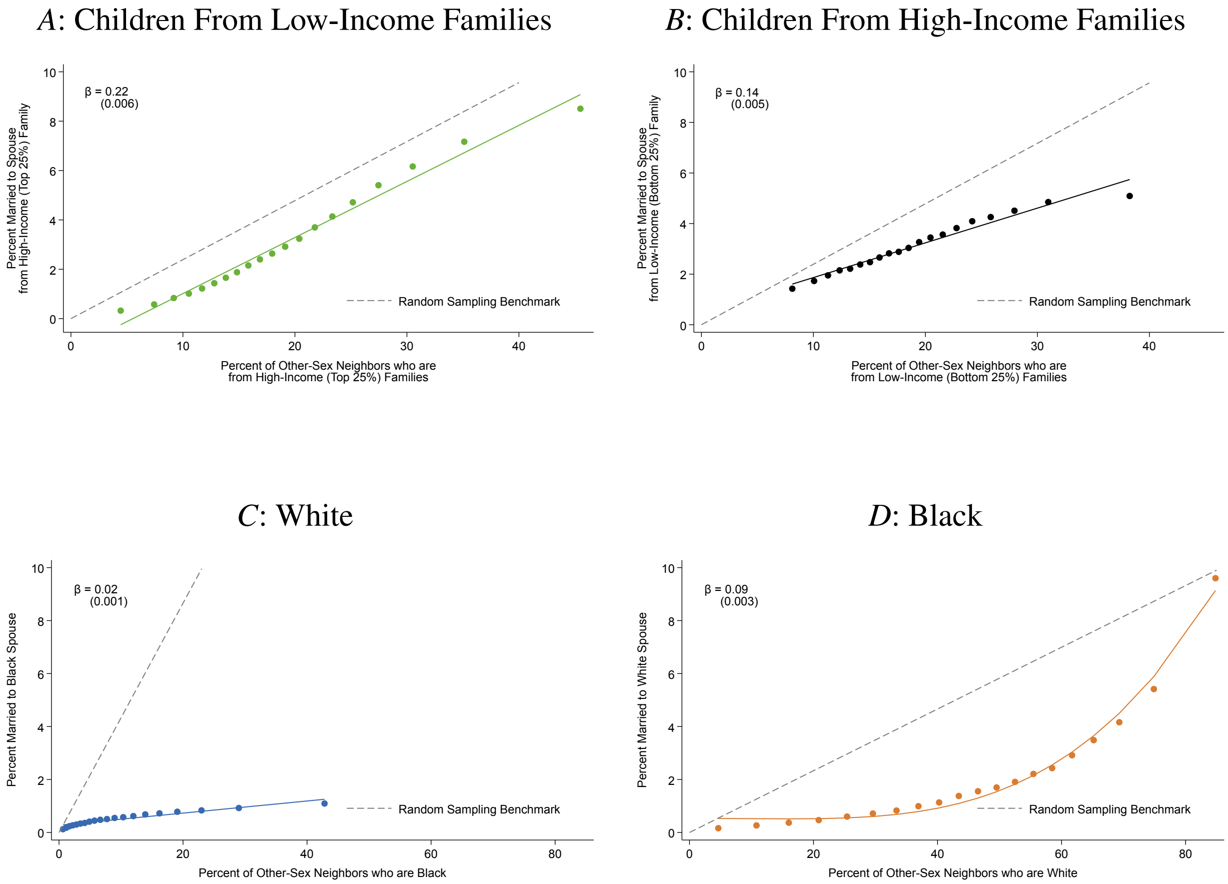


D: Black



Notes: This figure shows the distribution of neighborhood exposure to different groups in adulthood. In Panels *A* and *B*, we present the distribution of exposure that low and high parent income individuals have to opposite-sex peers who are themselves from low (Panel *A*) and high (Panel *B*) parent income families. In Panels *C* and *D*, we present the distribution of exposure that white and Black invididuals have to opposite-sex peers who are themselves white (Panel *C*) and Black (Panel *D*). In each panel, a given individual's exposure is determined using a time-weighted average of his/her addresses from age 18-27, imputed as described in Section II. We construct exposure using weights for geographic proximity and age proximity. We consider people in the nearest 50 Census tracts, which are weighted according to the decay weights shown in Figure II Panel *B*. We consider people who were in the same cohort as the individual or within the four cohorts younger and older than the individual. We use the sex-specific cohort weights shown in Figure A.2*B*. For details on the construction of the exposure measure, see Section III.C.

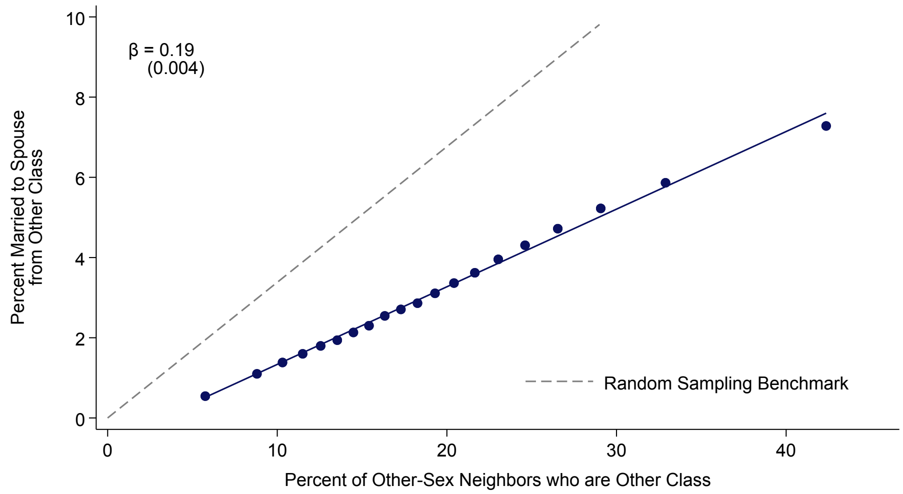
FIGURE IV: Association of Cross-Group Exposure and Cross-Group Marriage



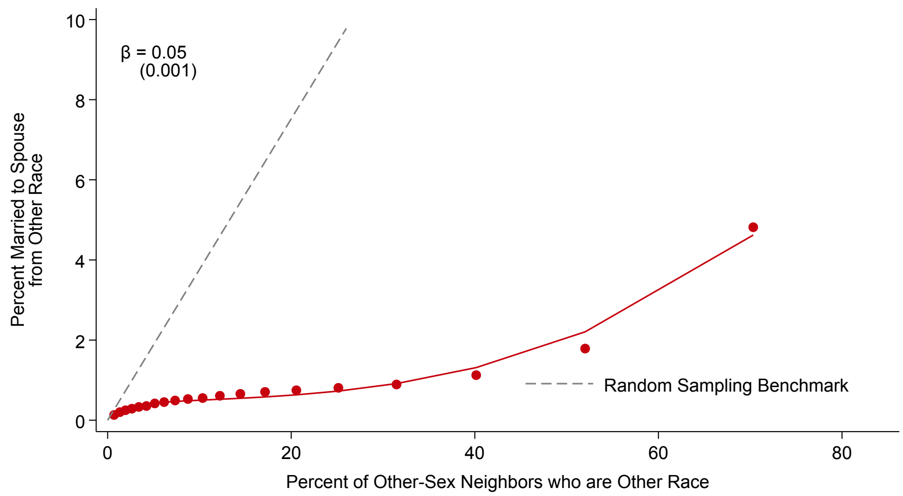
Notes: This figure presents binned scatter plots of the OLS relationship between cross-group exposure and cross-group marriage. In Panel *A* we show the relationship between the fraction of individuals from low parent income families who are married to a spouse from a high parent income family and the fraction of their opposite-sex peers who came from high-income families. In Panel *B* we show the analogous relationship, but using high parent income indviduals' exposure to opposite-sex low parent income peers and defining the outcome as having a spouse from a low-income familiy. In Panel *C* we show the relationship between the fraction of white individuals who are married to a Black spouse and exposure in adulthood to opposite-sex Black peers. In Panel *D* we show the analogous relationship, but using Black indviduals' exposure to opposite-sex white peers and defining the outcome as having a white spouse. Exposure is defined in adult neighborhoods between ages 18-27 as described in Section III.C. The distributions of these exposure measures are shown in Figure III. To construct each figure we first bin the exposure variable into 20 equally sized bins. Then we plot the mean exposure vs. the mean fraction married to the spouse type within each bin. The slope displayed in each panel is equal to the OLS slope from regressing the marriage outcome on the exposure variable in the micro level data, clustering the standard errors at the level of a person's age 23 county. We include a random sampling benchmark in the dashed line. The benchmark slope is equal to the group-specific marriage rate (e.g. the marriage rate for low parent income individuals in Panel *C*). The marriage rates used in the benchmarks can be found in Tables I and II for class and race, respectively.

FIGURE V: Impact of Race and Class Exposure on Spouse Characteristics

A: Rates of Interclass Marriage vs. Exposure to Other Class
Sample: Bottom 25% and Top 25%

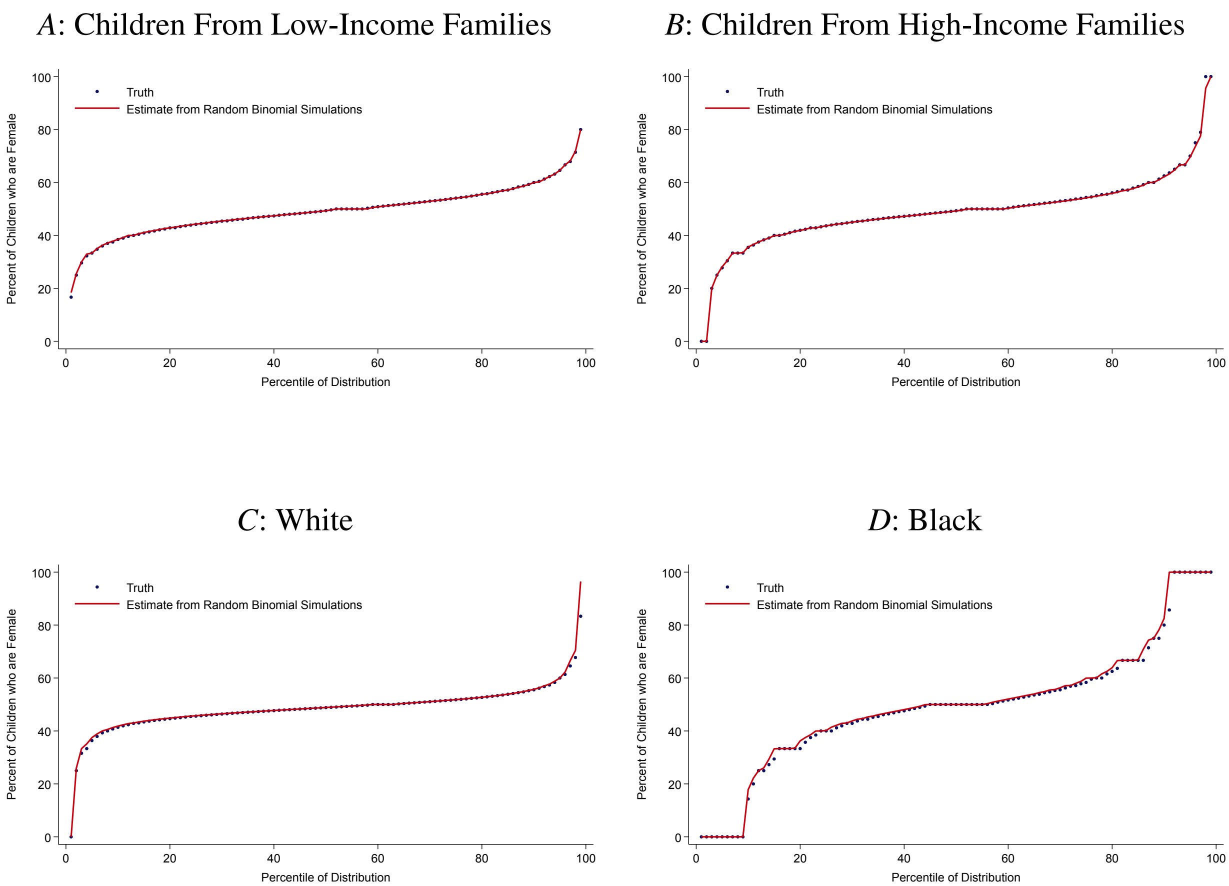


B: Rates of Interracial Marriage vs. Exposure to Other Race
Sample: White and Black



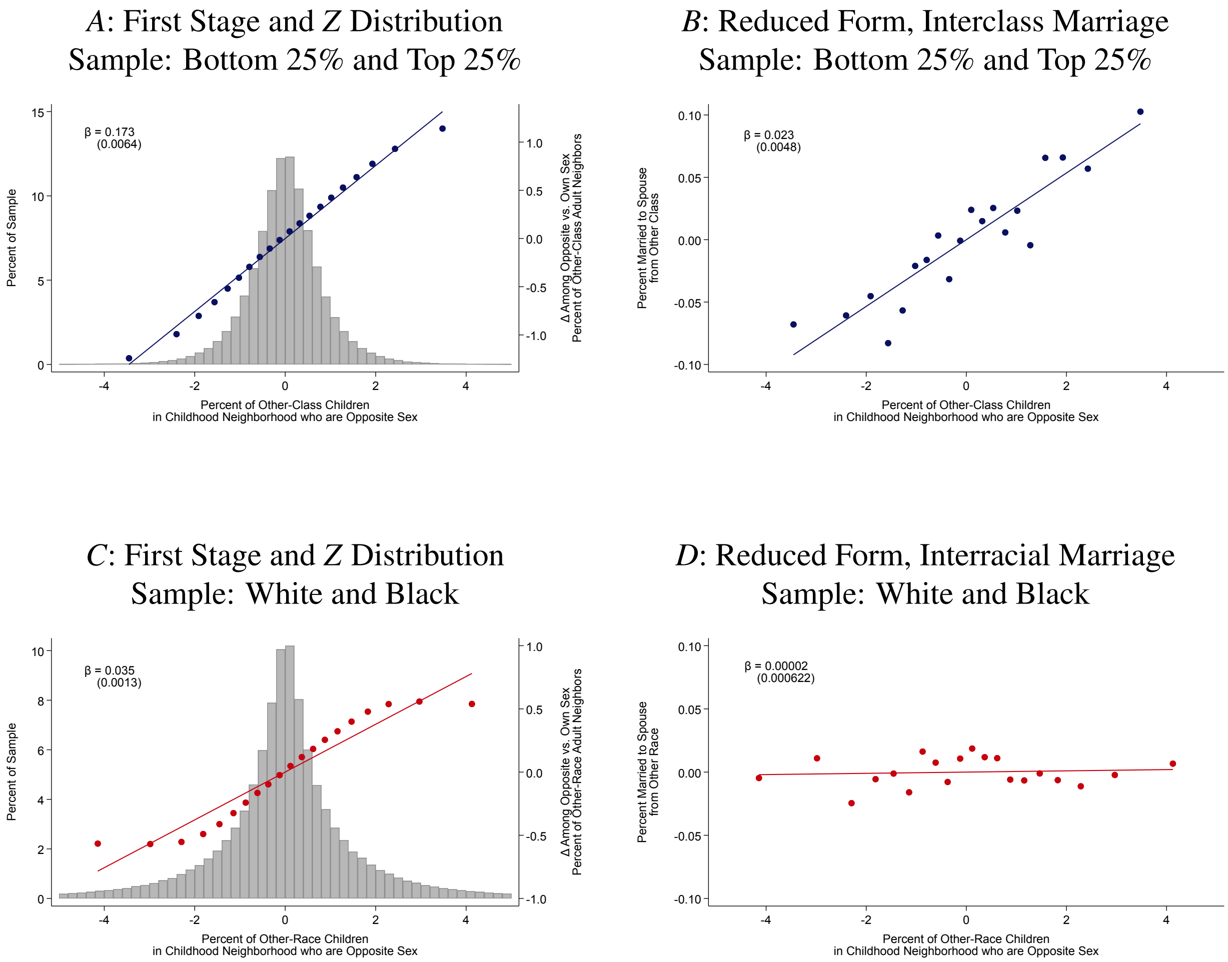
Notes: This figure presents binned scatter plots of the relationship between cross-group marriage and cross-group exposure, separately for class (Panel *A*) and race (Panel *B*). In Panel *A*, we stack the data used to create Figure IV Panels *A* and *B* and construct a single binned scatter plot. We begin with individuals who themselves come from low- or high-income families. Exposure for individuals from low-income families is exposure to opposite-sex peers from high-income families; for individuals from high-income families it is exposure to opposite-sex peers from low-income families. Similarly, for each group we define the marriage outcome as having a spouse from the other class group. Then we construct the binned scatter plot as in Figure IV; see figure notes for further details. In Panel *B*, we stack the data used to construct Figure IV Panels *C* and *D*. We begin with individuals who themselves white or Black and define exposure and marriage to the other race group. The slope displayed in each panel is equal to the OLS slope regressing the marriage outcome on the exposure variable in the micro level data, clustering the standard errors at the level of a person's age 23 county. We include a random sampling benchmark in each panel, with slope equal to the marriage rate for the class and race samples, respectively.

FIGURE VI: Distribution of Sex Ratio across Neighborhoods
Observed vs. Simulated



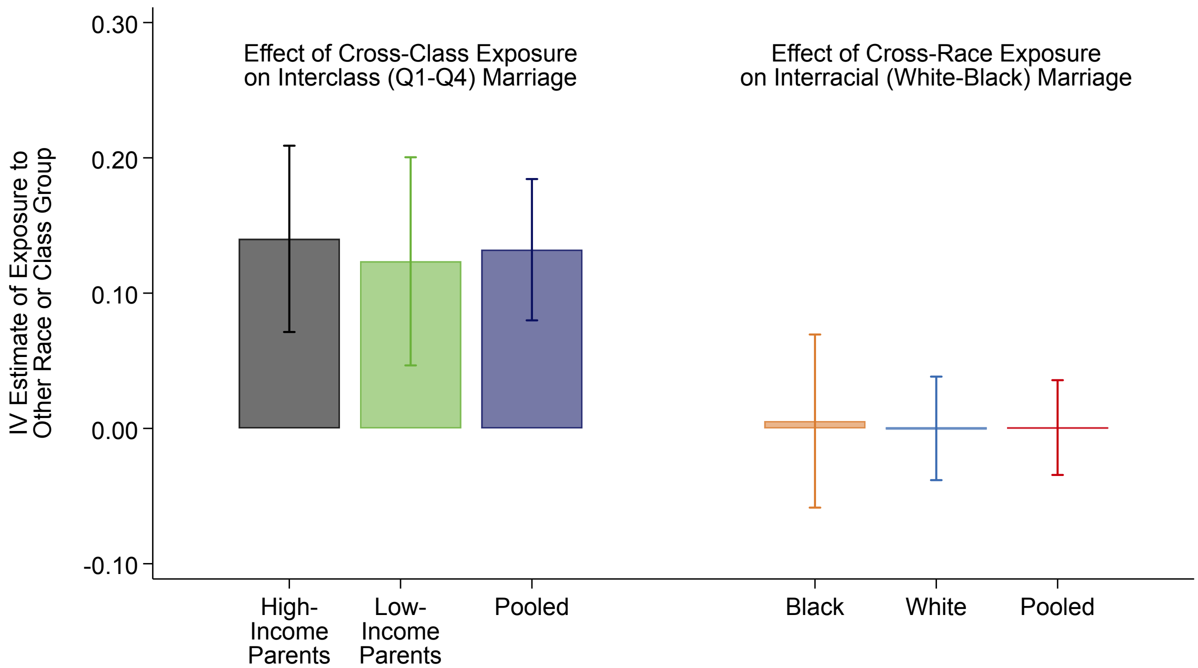
Notes: This figure assesses the extent to which observed sex ratio variation is consistent with random variation drawn from a binomial distribution, separately for individuals from low-and high-income families (Panels *A* and *B*) and for white and Black individuals (Panels *C* and *D*). We construct the sex ratio as the fraction of the individuals in each tract (and race or class group) who are female, restricting to people born between 1983-1987. We bin the tracts into percentiles and plot the mean fraction female in each bin against the percentile number. We then run random binomial simulations for comparison. For each run of the simulation, we draw the number of female individuals from a binomial distribution centered at the true sex ratio (approximately 49% female) with the number of draws equal to the total number of individuals in each tract and group. We construct percentiles of the sex ratio analogously to in the true data. We repeat that process 1,000 times. We take the mean within percentile across the 1,000 simulations and plot that on each figure.

FIGURE VII: Distribution of Instrument, First Stage, and Reduced Form



Notes: This figure presents the first stage and reduced form relationship between cross-group exposure and cross-group marriage for class (Panels *A* and *B*) and race (Panels *C* and *D*), respectively. The class results are estimated using individuals who are from families with parent income in the bottom or top quartile. The race results are estimated using white and Black individuals. The histogram in Panel *A* depicts the variation in the instrument - the fraction of other-class neighbors (e.g. from high-income families for individuals from low-income families) in an individual's childhood neighborhood (measured at age 18) who are the opposite sex. The distribution is residualized on class by sex by age 18 Census tract fixed effects and linear controls for the fraction of neighbors who are own-sex and the fraction of neighbors in the other class group, interacted with class and sex fixed effects. The first stage relationship between the fraction of other-sex neighbors in adulthood (measured over ages 18-27) who are other-class and the instrument is shown as a binned scatter plot. To construct the scatter, we follow Cattaneo et al. (2019). We first construct 20 equal sized bins of the instrument. We regress the the endogenous exposure measure on indicator variables for each bin, including the same controls as in the histogram. We plot the coefficients on the indicator variables against the mean of the instrument in the ventile bin, centering the plot at 0,0. We plot the first stage coefficient, estimated on running the specification with the same controls in the individual-level data and clustering standard errors at the age 23 county level. In Panel *B*, we plot the reduced form relationship as a binned scatter plot. We follow the same procedure to construct the figure as in the first stage. In Panels *C* and *D*, we construct analogous figures for cross-race exposure and interracial marriage.

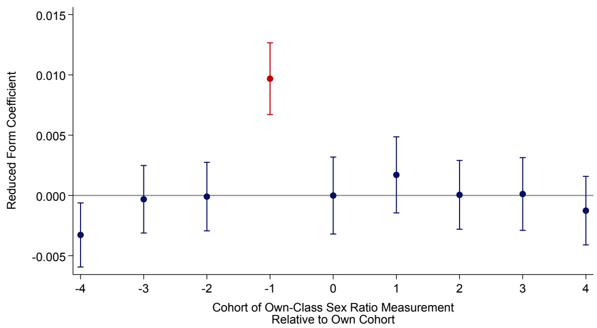
FIGURE VIII: IV Estimates of Cross-Group Exposure on Cross-Group Marriage



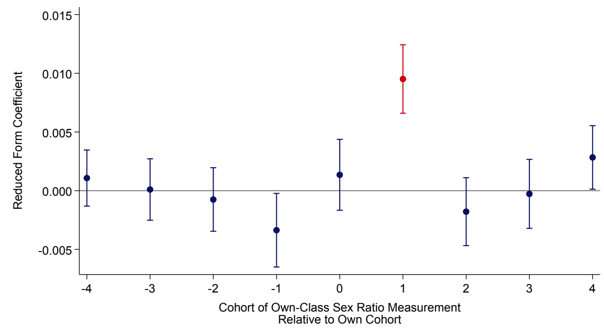
Notes: This figure summarizes the key IV results for interclass and interracial marriage. Results are presented separately for individuals from low- and high-income families as well as white and Black individuals as well as pooled for class and race, respectively. For details on the specifications procedure, see Section IV.B and the notes to Table VI. Additional robustness checks are shown in Table VI.

FIGURE IX: Outcome Based Placebo Test
Cohort Specific Sex Ratio Variation

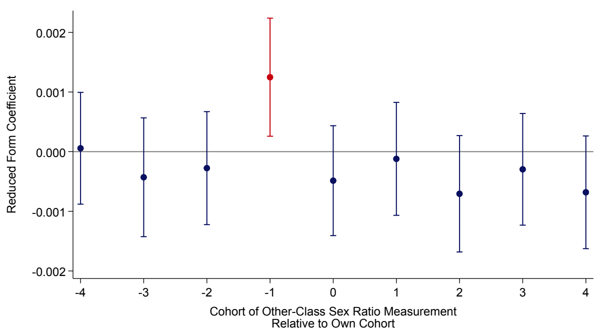
A: Own-Class Marriage to Spouse Born in Relative Cohort -1



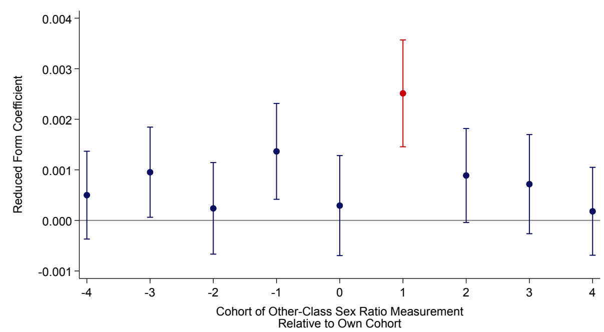
B: Own-Class Marriage to Spouse Born in Relative Cohort +1



C: Interclass Marriage to Spouse Born in Relative Cohort -1



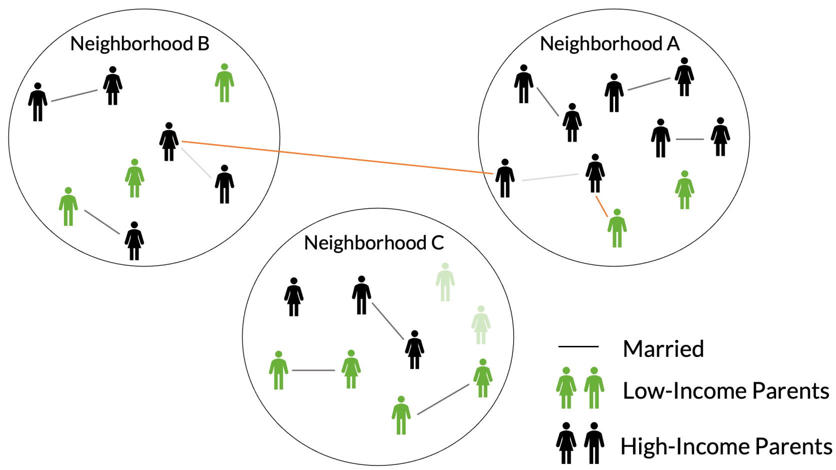
D: Interclass Marriage to Spouse Born in Relative Cohort +1



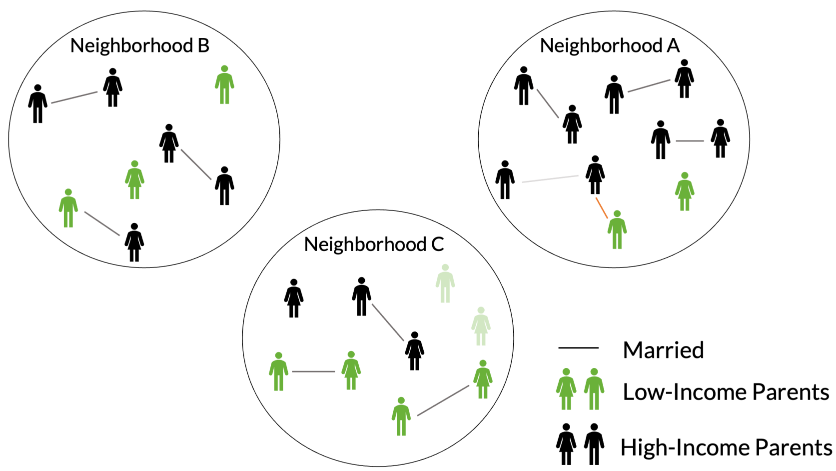
Notes: This figure presents a series of reduced form, outcome based placebo tests demonstrating the cohort specificity of the sex ratio variation. In each panel, the outcome variable is defined as having a spouse in a particular birth cohort by class group. In Panels *A* and *B*, the outcome is for an own-class spouse; in Panels *C* and *D*, an other-class spouse. In Panels *A* and *C*, the outcome is a spouse from the birth cohort one year younger than the individual; in Panels *B* and *D*, the cohort on year older. In each case, the outcome is regressed jointly on the 9 sex ratio instruments, defined as the sex ratio in each relative birth cohort. The baseline controls are included. See Section IV.B and the notes to Table VI for more details on the specification.

FIGURE X: Three Neighborhood Example of Spillover Effects

A: ★ Marries Person 1

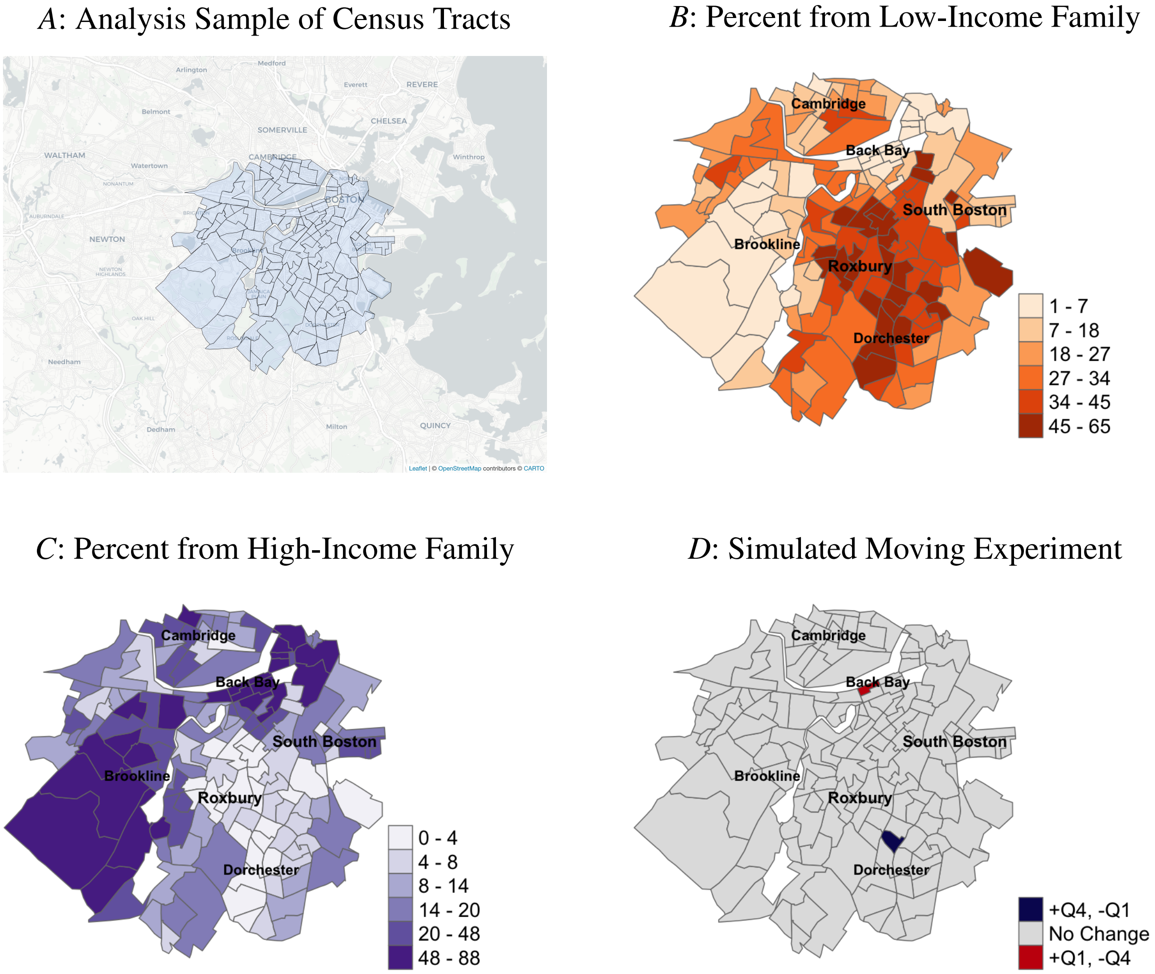


B: ★ Marries Person 2



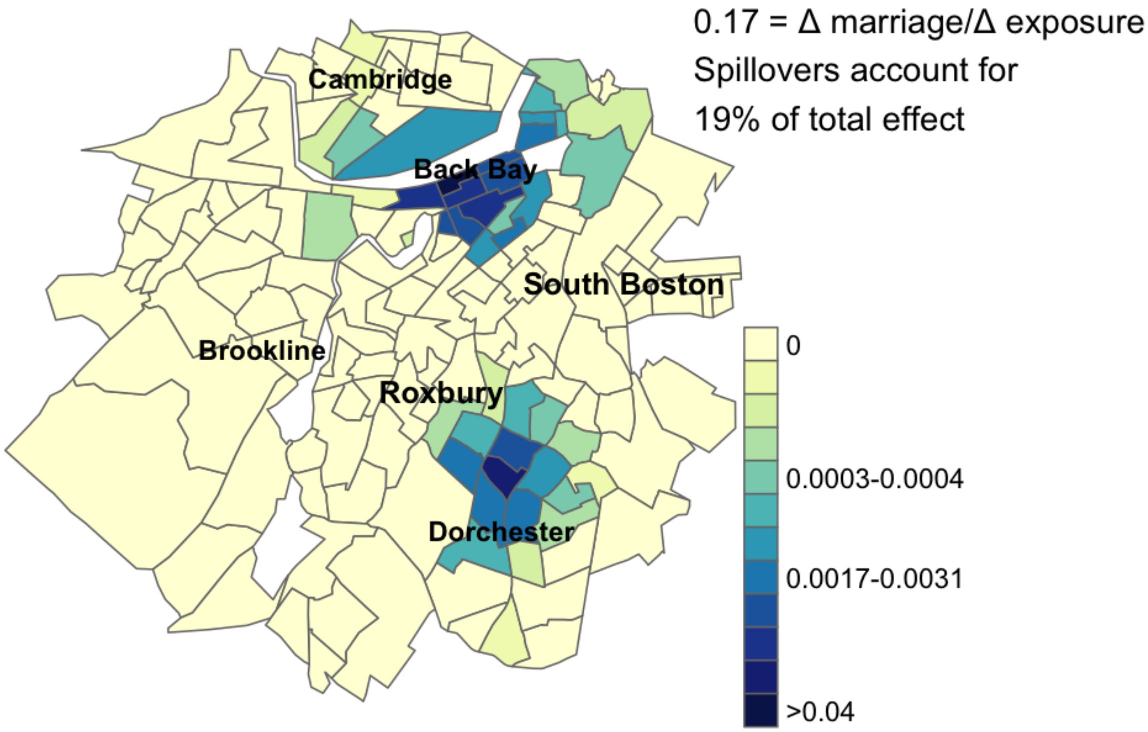
Notes: This figure presents a cartoon, three neighborhood example that illustrates the potential role of spillovers after a desegregation policy. In this example, there are two types of people, those from low- and high-income families, and the grey lines represent the initial marriages. Each panel depicts a possible counterfactual world after moving a resident from a low-income family from Neighborhood C to Neighborhood A, the reference neighborhood in this example. In the counterfactual worlds, the marriages change and the new pairings are represented by the orange lines. As a result of the move, in Panel *A*, a resident from a high-income family in Neighborhood B, denoted by the star, marries a a person from a high-income resident from Neighborhood C. In net, there is one new interclass marriage as a result of the policy and it is between an initial resident of Neighborhood A. In Panel *B*, the person denoted by the star marries a resident from a low-income family in Neighborhood C. In this counterfactual, there is still the new interclass marriage in Neighborhood A, but there is an additional interclass marriage from the star person. In this cartoon example, the spillovers account for 50% of the new interclass marriages.

FIGURE XI: Model Setup and Demopgrahics of Chosen Neighborhoods



Notes: This figure shows the region of Boston used to estimate the model described in Section V and the policy experiment described in Section V.C. In Panel *A* we show the 150 Census tracts used to estimate the model. In Panels *B* and *C* we plot the fraction of residents from low- and high-income families, respectively. Note that we adjust the overall proportion of residents from low- and high-income families across the entire area so that the shares are nationally representative. For more details, see Appendix C. In Panel *D*, we illustrate the policy experiment described in Section V.C. In it, two racially representative individuals from low-income families (one male and one female) are moved from a tract in Dorchester, shown in blue, to a tract in Backbay, shown in red, and are replaced with two analogous individuals from high-income families from Back Bay.

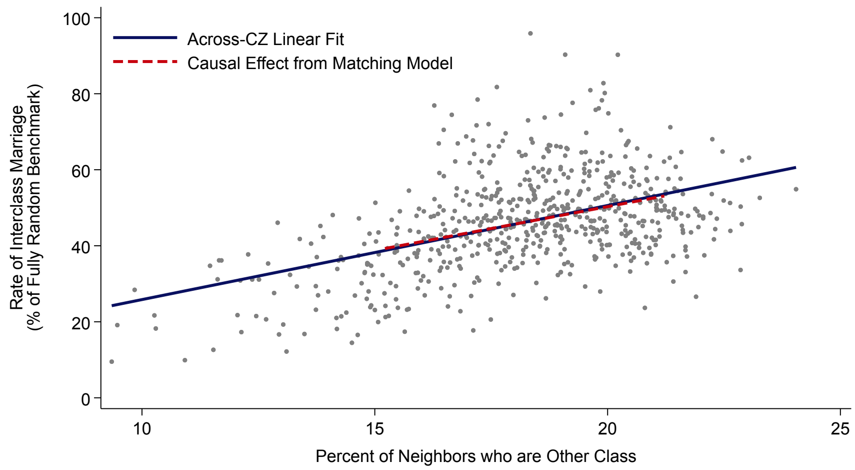
FIGURE XII: Impact of Policy Experiment on Interclass Marriage



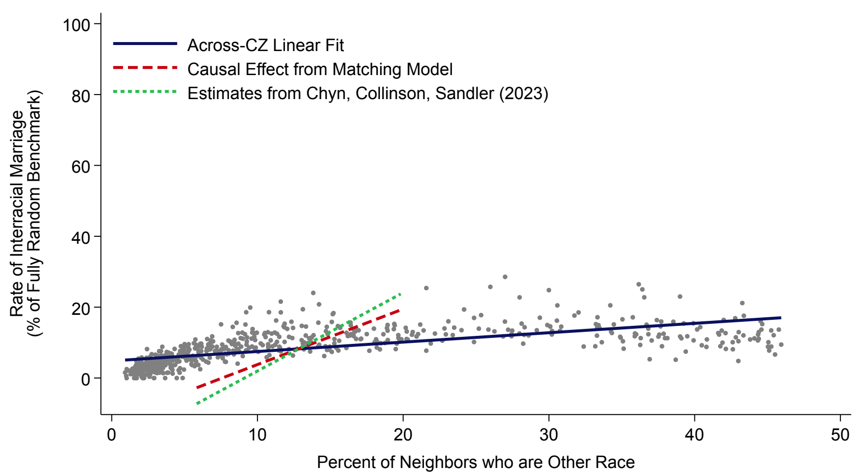
Notes: This figure presents the effect of the policy experiment shown in XI Panel *D* on interclass marriage. In the experiment, racially representative individuals from low-income families (one male and one female) are moved from a tract in Dorchester to a tract in Back Bay and they are replaced with racially representative inviduals from high-income families from Back Bay (one male and one female). The colors on the map represent the change in rates of interclass marriage predicted by the model as a result of the experiment. We report the change in the number of interclass marriages in the entire market and in the two focal tracts specifically. The difference in these measures represent spillovers are a result of the experiment.

FIGURE XIII: Commuting Zone Cross-Group Marriage vs. Cross-Group Exposure
Across Area Variation vs. Model Predictions from Policy Experiment

A: Rates of Interclass Marriage vs. Exposure to Other Class
Sample: Bottom 25% and Top 25%



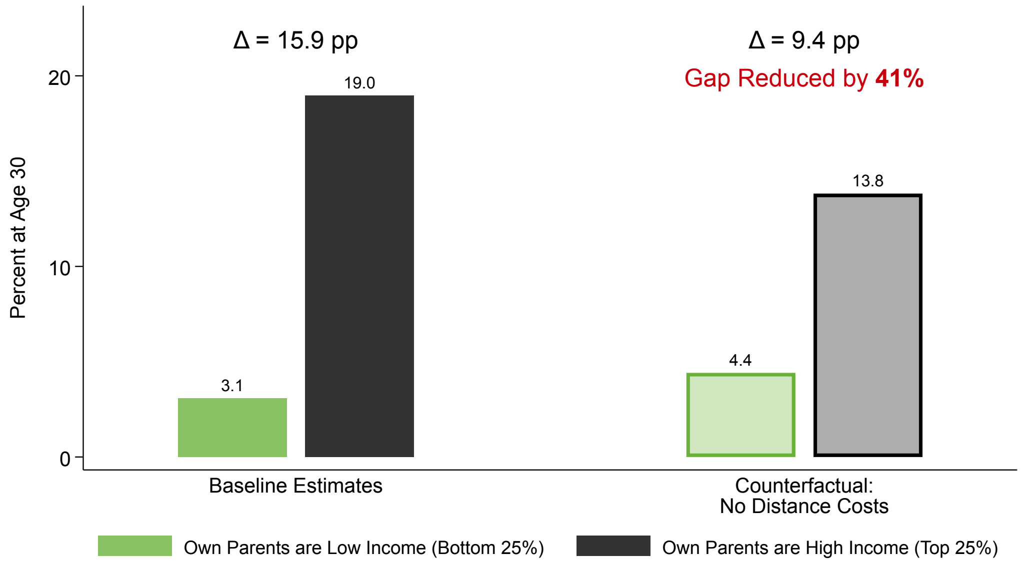
B: Rates of Interracial Marriage vs. Exposure to Other Race
Sample: White and Black



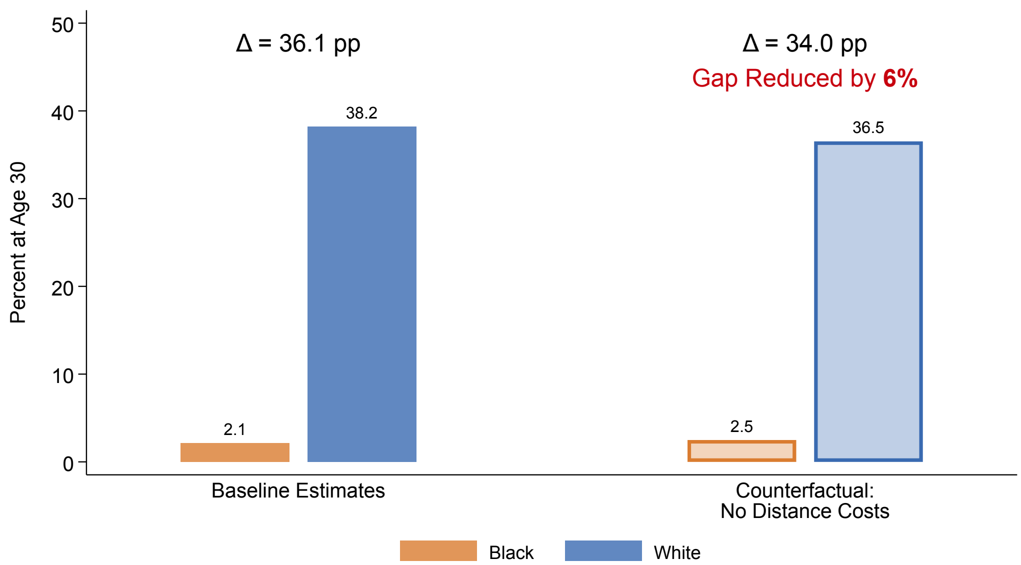
Notes: This figure plots the relationship between interclass (Panel *A*) and interracial (Panel *B*) marriage at the commuting zone level against exposure to neighbors of different class and race groups. This relationship is compared to estimates from the matching model developed in Section V. To construct the CZ measures, we assign individuals to the CZ they lived in at age 27. The exposure measure is the average of the exposure measure shown in Figure III across individuals in the CZ. The marriage rates are scaled relative to the fully random spouse benchmark shown in Figure I; see figure notes for additional details. The solid line shows the line of best fit across commuting zones, with even-weighting. The red, dashed lines in each panel show the implied slope from the counterfactual exercise described in SectionV.C and shown in XI Panel *D*. We compute the slope as the change in inter-group marriage as a result of the policy relative to the implied change in inter-group exposure. In Panel *B*, we additionally include a comparision to results from the Gatreaux Project (Chyn, Collinson and Sandler, 2023). They measure changes in exposure to white neighbors at the time of placement when the children were approximately 8 (Table 2 Column 4) and in 2019 when they were 38 on average (Table 5 Column 4). They also measure effects on interracial marriage (Table 6 Column 4). We construct the series by averaging the implied slope from the childhood and adulthood environment changes to better approximate our own exposure measure in young adulthood. The county level rates of interclass and interracial marriage will be available to download here.

FIGURE XIV: Impact of Search Costs on Marital Homophily by Class and Race

A: Married to Spouse with High-Income Parents



B: Married to White Spouse



Notes: This figure summarizes key patterns of marital homophily by class (parent income) and race as well as counterfactuals from a spatial model of the marriage market. The solid bars are repeated from Figure I; see figure notes for additional details. The hollow bars present a counterfactual from the model that removes any search costs across neighborhoods. In the baseline model described in Section V, individuals have preferences to marry a nearby neighbor, motivated by Figure II, that decay over the nearest 50 Census tracts. In the counterfactual, those preferences are set to zero, removing the effect of residential segregation on the pool of desirable spouses. Importantly, the group specific marriage rates (e.g. the overall fraction of individuals from low parent income families who are married) are held constant in the counterfactual. For additional details on the model setup and the counterfactual, see Section V.D.

Supplementary Appendix

A Exposure

As described in Section III, we define neighborhood using an individual's own tract at a given age and the nearest 50 census tracts. The neighbors are people who are within the four birth cohorts older and four birth cohorts younger than the individual. To construct our measures of exposure, we construct a weighted average over neighbors with separate weights for ordinal tract distance and relative birth cohort.

The weights are shown in Figure A.2. The relative cohort weights are sex-specific. We can define the weights formally as the share of marriages between individuals of a certain census tract distance or age-gap as the the number of marriages of those type, nationally, scaled by the number of marriages in the most common distance level or age-gap. As shown in Figure A.2 Panel *A*, we use separate distance weights for each tercile of the population density distribution, but we omit that from the notation in this section for simplicity. Let M^k denote the number of marriages between individuals k tracts away and $M^{m,s(i)}$ the number of marriages between individuals in relative birth cohort m for people who are sex s. We multiply the distance and age weights to get the following weights

$$\omega_{k,m,s} = \frac{M^k}{\max\limits_{k\in[0,50]}\{M^k\}} \frac{M^{m,s}}{\max\limits_{m\in[-4,4]}\{M^{s,m}\}}.$$

Define $l_\alpha(i)$ as the census tract for person i at age α. Let $n_{g,l_\alpha(i),k,b(i)+m}$ represent the raw number of individuals who are group g in the kth nearest tract to tract $l(i,\alpha)$ and who are born in birth cohort $b(i)+m$. We define $n_{l(i,\alpha),k,b(i)+m}$ as the total number of individuals in that tract neighbor and relative cohort cell. Combining these raw counts with the weights gives the following weighted sums

$$N_{i,g}^{\alpha} = \sum_{k=0}^{50}\sum_{m=-4}^{m=4} \omega_{k,m,s(i)} n_{g,l(i,\alpha),k,b(i)+m}$$
$$N_{i}^{\alpha} = \sum_{k=0}^{50}\sum_{m=-4}^{m=4} \omega_{k,m,s(i)} n_{l(i,\alpha),k,b(i)+m}$$

for the number of type g neighbors and total neighbors for person i in their age a neighborhood.

We define exposure, at a given age, α, as

$$e^{\alpha}_{i,g} = \frac{N^{\alpha}_{i,g}}{N^{\alpha}_{i}}.$$

For much of our analysis, we define one's adult neighborhood over ages 18-27, which we refer to with superscript a, for simplicity. In this case, as in the exposure distributions in Figure III, we even-weight over the ages to measure aggregate exposure over this decadal period. We can write this as

$$e^{a}_{i,g} = \frac{1}{10}\sum_{j=18}^{27}\frac{N^{j}_{i,g}}{N^{j}_{i}}.$$

Note that in the exposure distributions histograms, the relative weights, $\omega_{k,m,s}$, are sex-specific, but the raw counts contain all individuals, regardless of sex. When defining the endogenous treatment variable, $T_{i,g}$, in Section IV.B, we consider the difference in exposure to type g among own-sex and opposite-sex neighbors. In this case, we define the sex-specific weighted counts at age α as the following for own- and opposite-sex, respectively, as

$$\begin{aligned}
N^{\alpha}_{i,s(i),g} &= \sum_{k=0}^{50}\sum_{m=-4}^{m=4}\omega_{k,m,s(i)}n_{g,s(i),l(i,\alpha),k,b(i)+m} \\
N^{\alpha}_{i,s(i)} &= \sum_{k=0}^{50}\sum_{m=-4}^{m=4}\omega_{k,m,s(i)}n_{s(i),l(i,\alpha),k,b(i)+m} \\
N^{\alpha}_{i,-s(i),g} &= \sum_{k=0}^{50}\sum_{m=-4}^{m=4}\omega_{k,m,s(i)}n_{g,-s(i),l(i,\alpha),k,b(i)+m} \\
N^{\alpha}_{i,-s(i)} &= \sum_{k=0}^{50}\sum_{m=-4}^{m=4}\omega_{k,m,s(i)}n_{-s(i),l(i,\alpha),k,b(i)+m}.
\end{aligned}$$

Using these definitions, we can write the endogenous exposure measure as the average over ages 18-27 as

$$T_{i,g} = \frac{N^{a}_{i,-s,g}}{N^{a}_{i,-s}} - \frac{N^{a}_{i,s,g}}{N^{a}_{i,s}} = \frac{1}{10}\sum_{j=18}^{27}\left(\frac{N^{j}_{i,-s(i),g}}{N^{j}_{i,-s(i)}} - \frac{N^{j}_{i,s(i),g}}{N^{j}_{i,s(i)}}\right).$$

B Relationship Between Treatment and Instrument

Define the fraction of childhood neighbors who are from group g as

$$P^c_{i,g} \equiv \frac{N^c_{i,g}}{N^c_i}$$

and the fraction who are from sex s as

$$P^c_{i,s} \equiv \frac{N^c_{i,s}}{N^c_i}.$$

so that P^c_g is the fraction of neighbors during childhood who are from group g and P^c_s is the fraction who are from sex s. Also, define the childhood neighborhood version of market tightness is

$$T^c_{i,g} = \frac{N^c_{i,-s,g}}{N^c_{i,-s}} - \frac{N^c_{i,s,g}}{N^c_{i,s}}.$$

Then, using Bayes' rule, we can express market tightness as a function of the instrument and the marginal group and sex shares by

$$T^c_{i,g} = Z_{i,g}\frac{P^c_{i,g}}{P^c_{i,s}} - (1 - Z_{i,g})\frac{P^c_{i,g}}{1 - P^c_{i,s}} = Z_{i,g}\left(\frac{P^c_{i,g}}{P^c_{i,s}} + \frac{P^c_{i,g}}{1 - P^c_{i,s}}\right) - \frac{P^c_{i,g}}{1 - P^c_{i,s}}. \tag{B1}$$

The above formulation shows that, conditional on $P^c_{i,g}$ and $P^c_{i,s}$, is a linear function of $Z_{i,g}$. We control for $P^c_{i,g}$ and $P^c_{i,s}$ in our IV estimates to generate a linear relationship between treatment and instrument (see Figure VII). In practice, our treatment variable $T_{i,g}$ will differ from $T^c_{i,g}$ because some individuals migrate out of childhood neighborhoods. As a result, our actual first stage coefficient will be smaller than $\left(\frac{P^c_{i,g}}{P^c_{i,s}} + \frac{P^c_{i,g}}{1-P^c_{i,s}}\right)$, which is what we would get if everybody remained in their childhood neighborhood.

C Constructing the Sample of Tracts and Types

The types in the model are defined at the intersection of the following characteristics

$$j,k \in \{\text{Parent Income Quartile} \times \text{Race} \times \text{Adult Census Tract} \times \text{Unobserved Trait}\}.$$

Including Census tracts in the type space adds substantial computational complexity to the model. As we discussed in Section V.B, μ would have a dimension of more than five trillion if the model were solved on the entire U.S. Instead, we solve the model on a sample of 150 census tracts so that the dimension of μ is approximately 23 million. We choose the sample of tracts shown in Figure XI Panel *A* and construct the sample to be nationally representative using the procedure described below. The nationally representative sample is important because the moments the model is being matched to are estimated nationally. Because the model does not allow preferences to be vary by geographic area, the chosen sample matters only because it determines the topography of segregation, i.e., who lives where.

Our model is estimated outside of the Census data environment. As a result, we have to estimate tract level counts in each parent income quartile and race cell since there are no external data with this information. To do so, we begin with the data at the census tract and race level published in Chetty et al. (2018). We restrict the sample to tracts with at least 100 people living in them in the 2000 Census. Within each tract and race cell, these data also include the fraction of individuals who have parents with below median parental income and the mean parent income percentile.

We predict the share of individuals in a tract and race who are from families in a given parent income quartile by regressing

$$y_{t,q,r} = \alpha_r + \beta_1 p_t + \beta_2 s_t + \beta_3 p_t \times s_t + \varepsilon_{t,q,r}.$$

Where $y_{t,q,r}$ is the share of individual in tract t and race r who came from families in parent income quartile q, α_r are race group fixed effects, p_t is the mean parent income percentile in tract t, and s_t is the share of individuals who have below median income parents. We use this regression to predict $\hat{y}_{t,1,r}$ and $\hat{y}_{t,4,r}$—we attain R-squared values above 0.9 in both cases. We define $\hat{y}_{t,2,r}$ using the fact that $\hat{y}_{t,1,r} + \hat{y}_{t,2,r} = s_t$ and $\hat{y}_{t,3,r}$ using the fact that $\hat{y}_{t,3,r} + \hat{y}_{t,4,r} = 1 - s_t$. The $\hat{y}$ values then get rescaled so that the sample of tracts in Figure XI Panel *A* is nationally representative by race and parent income quartile.[1] Once we have counts in each tract by race by parent income quartile cell, we simulate sex ratio variation using random draws from a binomial distribution.

The next step is to assign the share of each tract, t, parent income quartile, q, and race, r, that has the unobserved trait. We do so using the logistic function

$$u_{t,q,r} = \frac{e^{\Omega_{t,q,r}}}{1 + e^{\Omega_{t,q,r}}}$$

to constrain the share to be between zero and one. The $\kappa_{t,q,r}$ is defined as

$$\Omega_{t,q,r} = \omega_q^1 + \omega_r^2 + (\omega_q^3 + \omega_r^4) p_t$$

[1]The rescaling is done by first multiplying the counts in each tract by parent income quartile by race cell by one constant per race group and then one constant by parent income quartile to ensure that 25% of individuals belong to each quartile.

so that the share with the unobserved trait is allowed to vary by race, class, and the type of neighborhood, as measured by the average income of parents, p_t. The ω parameters are estimated.

D Model Estimation

Having shown how the $\{m, w\}$ values are computed in Appendix C, we now turn to computing preferences. The equation for marital surplus is given by Equation 9 in Section V.A. The distance cost portion of surplus is defined by

$$f(j,k;\zeta) = \sum_{h \in \{j,k\}} \left[\mathbb{1}\{n(j,k) \leq 50\} \times (\zeta^1_{r(h)} + \zeta^2_{c(h)}) \left(\frac{1}{n(j,k)+1} \right)^{\kappa_{c(h),r(h)}} \right]$$

where $n(j,k)$ is a function that returns the neighbor number of tract k from the perspective of j, e.g., $n(j,k) = 10$ if tract k is the 10th closest tract to tract j. The parameters $\zeta^1_{r(h)}$ and $\zeta^2_{c(h)}$ determine the race and class specific premium to marrying a spouse who is living within 50 census tracts. The

$$\kappa_{c(h),r(h)} = \frac{\phi e^{\zeta^3_{r(h)} + \zeta^4_{c(h)}}}{1 + e^{\zeta^3_{r(h)} + \zeta^4_{c(h)}}}$$

parameter controls the steepness of the distance cost function, i.e., within 50 census tracts, how much do people of a given race and class group prefer to marry somebody in a nearer vs. further tract.[2] We estimate the ζ parameters as part of our simulated method of moments procedure.

Once $\gamma_{j,k}$ is known, we can solve the equilibrium matches, μ using the system

$$m_j = \mu_{j,0} + \sum_{k=1}^{T} \sqrt{e^{\gamma_{j,k}} \mu_{j,0} \mu_{0,k}}$$
$$w_k = \mu_{0,k} + \sum_{j=1}^{T} \sqrt{e^{\gamma_{j,k}} \mu_{j,0} \mu_{0,k}}.$$

In a sample of 150 tracts, the total number of types in the model, T, is 4800. This is a system of $2T$ equations and $2T$ unknowns. Solving using the analytic Jacobian is slow and computationally expensive because the Jacobian has $4T^2$ elements which is almost 100 million in this setting. Instead, we can turn this into a system of quadratic equations if we change variables such that

[2] The ϕ parameter puts a maximum value on how large this exponent can be. We set ϕ to 0.15 for the baseline estimates.

$\pi = \sqrt{\mu}$. Define $\nu_{m,j} \equiv \sum_{k=1}^{T} \pi_{0,k}\sqrt{e^{\gamma_{j,k}}}$ so that the male equilibrium equations become

$$m_j = \pi_{j,0}^2 + \pi_{j,0}\nu_{m,j}.$$

We can apply the quadratic equation to solve

$$\pi_{j,0} = \frac{-\nu_{m,j} + \sqrt{\nu_{m,j}^2 + 4m_j}}{2}$$

and the analogous equations for the female equilibrium

$$\pi_{0,k} = \frac{-\nu_{w,k} + \sqrt{\nu_{w,k}^2 + 4w_k}}{2}.$$

This system of equations can be solved with a fixed point algorithm. We implement the Anderson (1965) algorithm with JAX, which does fast floating point operations on graphical processing units (Bradbury et al., 2018). After this is solved, we use $\mu = \pi^2$ to obtain μ.

Once we have μ then we can compute the partial equilibrium moments, $\hat{\psi}_s(\theta)$, at the current value of θ. We can then continue to adjust θ to minimize the SMM objective function shown in 10. Our baseline estimates use a diagonal W that places half of the weight on the national marriage outcomes and half on the OLS and IV exposure effects. The remaining weight is put on the OLS and IV exposure effects in approximate proportion to their precision.[3] We also plan to include estimates where W is based off the covariance matrix of $\hat{\psi}_d$ in a future version of this paper. The reason we do not do this for our baseline estimates is because the national marriage outcomes are estimated with extremely high precision since our data cover the full population. This means the exposure effects would end up with approximately zero weight in the objective function.

[3]Our OLS estimates are ≈ 10 times as precise as the IV estimates. As a fraction of the total weight on the exposure effects, we put 90% of the weight on the OLS estimates and 10% of the weight on the IV estimates.

TABLE A.1: Cohabitation Outcomes by Class Relative to Random Matching Benchmark

			Fraction with Partner from Parent Income Group at Age 30				
Parent Income Group		Fraction Cohabiting (1)	Bottom 25% (2)	Quartile 2 (3)	Quartile 3 (4)	Top 25% (5)	Out of Sample (6)
Bottom 25%	Truth	0.427	0.097	0.092	0.078	0.051	0.109
	Fully Random	0.494	0.103	0.103	0.103	0.103	0.082
	Random Partner	0.427	0.077	0.084	0.096	0.099	0.071
Quartile 2	Truth	0.467	0.082	0.105	0.111	0.080	0.088
	Fully Random	0.494	0.103	0.103	0.103	0.103	0.082
	Random Partner	0.467	0.084	0.092	0.104	0.109	0.078
Quartile 3	Truth	0.530	0.065	0.105	0.148	0.141	0.071
	Fully Random	0.494	0.103	0.103	0.103	0.103	0.082
	Random Partner	0.530	0.096	0.104	0.119	0.123	0.088
Top 25%	Truth	0.551	0.042	0.076	0.138	0.235	0.061
	Fully Random	0.494	0.103	0.103	0.103	0.103	0.082
	Random Partner	0.551	0.099	0.109	0.123	0.128	0.092

Notes: This table presents observed rates of interclass cohabitation, measured at age 30, and two benchmarks. Cohabitation is defined using data from the ACS. Concretely, we take individuals in the main analysis sample who receive the ACS in the year they are 30. Individuals are cohabiting if they respond to the survey as the primary respondent and list a married spouse or a non-married partner in their household as an additional respondent or if they are listed as the spouse or partner of the primary respondent. The benchmarks are defined analogously to those in Table III, using cohabitation rather than marriage. The fully random benchmark is the overall cohabitation rate multiplied by the fraction of the population from each class group. The random partner benchmark is equal to the actual cohabitation rate for each race group multiplied by the fraction of the cohabiting population that comes from each class group.

TABLE A.2: Cohabitation Outcomes by Race Relative to Random Matching Benchmark

			Fraction with Partner from Race Group at Age 30						
Race Group		Fraction Cohabiting (1)	White (2)	Black (3)	Asian (4)	Hispanic (5)	AIAN (6)	Other (7)	Out of Sample (8)
White	Truth	0.568	0.503	0.009	0.008	0.030	0.003	0.010	0.006
	Fully Random	0.498	0.317	0.063	0.018	0.071	0.004	0.013	0.011
	Random Partner	0.568	0.414	0.040	0.015	0.070	0.004	0.012	0.013
Black	Truth	0.276	0.044	0.191	0.003	0.018	0.001	0.009	0.010
	Fully Random	0.498	0.317	0.063	0.018	0.071	0.004	0.013	0.011
	Random Partner	0.276	0.201	0.019	0.007	0.034	0.002	0.006	0.006
Asian	Truth	0.360	0.113	0.008	0.185	0.024	0.001	0.021	0.009
	Fully Random	0.498	0.317	0.063	0.018	0.071	0.004	0.013	0.011
	Random Partner	0.360	0.262	0.025	0.010	0.045	0.002	0.008	0.008
Hispanic	Truth	0.430	0.123	0.013	0.007	0.241	0.002	0.009	0.035
	Fully Random	0.498	0.317	0.063	0.018	0.071	0.004	0.013	0.011
	Random Partner	0.430	0.313	0.030	0.011	0.053	0.003	0.009	0.010
AIAN	Truth	0.426	0.194	0.022	0.008	0.044	0.130	0.015	0.012
	Fully Random	0.498	0.317	0.063	0.018	0.071	0.004	0.013	0.011
	Random Partner	0.426	0.310	0.030	0.011	0.053	0.003	0.009	0.009
Other	Truth	0.428	0.241	0.040	0.030	0.048	0.005	0.058	0.007
	Fully Random	0.498	0.317	0.063	0.018	0.071	0.004	0.013	0.011
	Random Partner	0.428	0.312	0.030	0.011	0.053	0.003	0.009	0.010

Notes: This table presents observed rates of interracial cohabitation, measured at age 30, and two benchmarks. Cohabitation is defined using data from the ACS. Concretely, we take individuals in the main analysis sample who receive the ACS in the year they are 30. Individuals are cohabiting if they respond to the survey as the primary respondent and list a married spouse or a non-married partner in their household as an additional respondent or if they are listed as the spouse or partner of the primary respondent. The benchmarks are defined analogously to those in Table IV, using cohabitation rather than marriage. The fully random benchmark is the overall cohabitation rate multiplied by the fraction of the population from each race group. The random partner benchmark is equal to the actual cohabitation rate for each race group multiplied by the fraction of the cohabiting population that comes from each race group.

TABLE A.3: Marriage Outcomes by Class Relative to Random Matching Benchmark
Sample: 1982 Birth Cohort

			Fraction with Spouse from Parent Income Group at Age 37				
Parent Income Group		Fraction Married (1)	Bottom 25% (2)	Quartile 2 (3)	Quartile 3 (4)	Top 25% (5)	Out of Sample (6)
Bottom 25%	Truth	0.297	0.053	0.056	0.053	0.039	0.097
	Fully Random	0.451	0.088	0.088	0.088	0.088	0.101
	Random Spouse	0.297	0.038	0.051	0.065	0.076	0.066
Quartile 2	Truth	0.402	0.056	0.078	0.090	0.077	0.100
	Fully Random	0.451	0.088	0.088	0.088	0.088	0.101
	Random Spouse	0.402	0.051	0.069	0.088	0.103	0.090
Quartile 3	Truth	0.511	0.053	0.090	0.131	0.135	0.101
	Fully Random	0.451	0.088	0.088	0.088	0.088	0.101
	Random Spouse	0.511	0.065	0.088	0.113	0.131	0.114
Top 25%	Truth	0.594	0.040	0.077	0.136	0.236	0.105
	Fully Random	0.451	0.088	0.088	0.088	0.088	0.101
	Random Spouse	0.594	0.076	0.103	0.131	0.152	0.133

Notes: This table presents statistics on rates of interclass marriage measured at age 37. The sample contains individuals in the 1982 birth cohort, the oldest in the main analysis sample, who are 37 in 2019, the last year available in the tax data. For details on construction, see the notes to Table III.

TABLE A.4: Marriage Outcomes by Race Relative to Random Matching Benchmark
Sample: 1982 Birth Cohort

			Fraction with Spouse from Race Group at Age 37						
Race Group		Fraction Married (1)	White (2)	Black (3)	Asian (4)	Hispanic (5)	AIAN (6)	Other (7)	Out of Sample (8)
White	Truth	0.546	0.482	0.006	0.008	0.025	0.002	0.008	0.014
	Fully Random	0.463	0.292	0.060	0.015	0.060	0.004	0.011	0.020
	Random Spouse	0.546	0.406	0.025	0.020	0.057	0.003	0.011	0.024
Black	Truth	0.161	0.028	0.104	0.002	0.010	0.000	0.005	0.012
	Fully Random	0.463	0.292	0.060	0.015	0.060	0.004	0.011	0.020
	Random Spouse	0.161	0.120	0.007	0.006	0.017	0.001	0.003	0.007
Asian	Truth	0.513	0.123	0.009	0.267	0.027	0.001	0.024	0.064
	Fully Random	0.463	0.292	0.060	0.015	0.060	0.004	0.011	0.020
	Random Spouse	0.513	0.382	0.023	0.019	0.053	0.003	0.011	0.022
Hispanic	Truth	0.371	0.112	0.010	0.007	0.187	0.002	0.008	0.046
	Fully Random	0.463	0.292	0.060	0.015	0.060	0.004	0.011	0.020
	Random Spouse	0.371	0.276	0.017	0.014	0.039	0.002	0.008	0.016
AIAN	Truth	0.304	0.154	0.009	0.004	0.025	0.087	0.012	0.013
	Fully Random	0.463	0.292	0.060	0.015	0.060	0.004	0.011	0.020
	Random Spouse	0.304	0.226	0.014	0.011	0.032	0.002	0.006	0.013
Other	Truth	0.408	0.211	0.028	0.035	0.042	0.004	0.062	0.026
	Fully Random	0.463	0.292	0.060	0.015	0.060	0.004	0.011	0.020
	Random Spouse	0.408	0.303	0.019	0.015	0.043	0.002	0.008	0.018

Notes: This table presents statistics on rates of interracial marriage measured at age 37. The sample contains individuals in the 1982 birth cohort, the oldest in the main analysis sample, who are 37 in 2019, the last year available in the tax data. For details on construction, see the notes to Table IV.

TABLE A.5: Marriage Outcomes by Class and Sex

			Fraction with Spouse from Parent Income Group at Age 30				
Parent Income Group		Fraction Married (1)	Bottom 25% (2)	Quartile 2 (3)	Quartile 3 (4)	Top 25% (5)	Out of Sample (6)
Bottom 25%	Pooled	0.239	0.050	0.052	0.047	0.031	0.059
	Male	0.217	0.048	0.050	0.046	0.029	0.043
	Female	0.262	0.052	0.053	0.048	0.033	0.076
Quartile 2	Pooled	0.317	0.052	0.070	0.079	0.059	0.056
	Male	0.286	0.050	0.067	0.076	0.055	0.038
	Female	0.349	0.055	0.074	0.082	0.063	0.075
Quartile 3	Pooled	0.405	0.048	0.080	0.117	0.111	0.050
	Male	0.363	0.044	0.074	0.110	0.104	0.031
	Female	0.449	0.051	0.086	0.124	0.118	0.070
Top 25%	Pooled	0.437	0.031	0.060	0.111	0.190	0.045
	Male	0.388	0.029	0.054	0.102	0.174	0.028
	Female	0.489	0.034	0.066	0.121	0.205	0.063

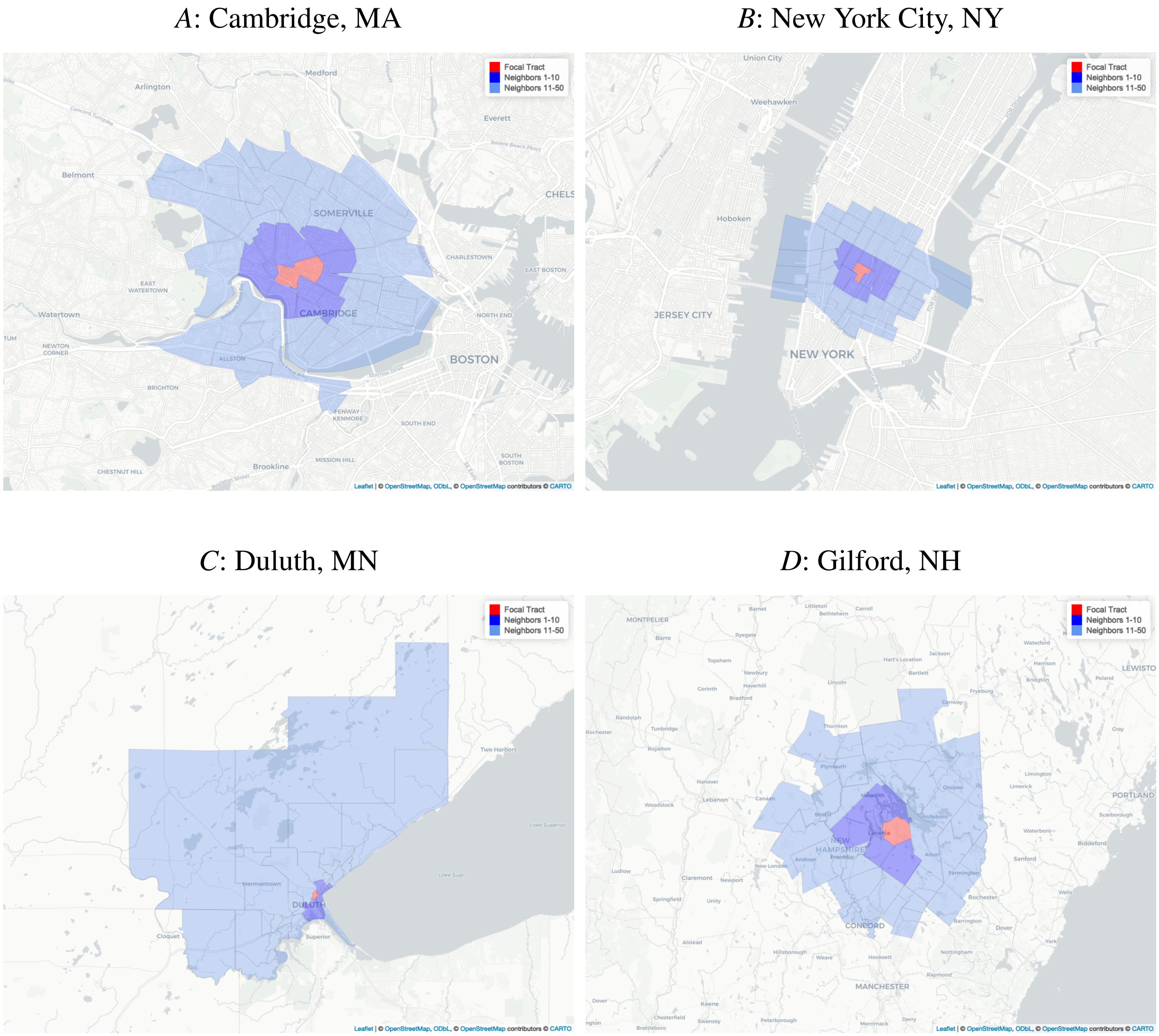
Notes: This table presents statistics on rates of interclass marriage separately by sex. For details on construction, see the notes to Table III.

TABLE A.6: Marriage Outcomes by Race and Sex

			Fraction with Spouse from Race Group at Age 30						
		Fraction Married	White	Black	Asian	Hispanic	AIAN	Other	Out of Sample
Race Group		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
White	Pooled	0.433	0.382	0.005	0.005	0.020	0.002	0.007	0.012
	Male	0.386	0.339	0.003	0.006	0.019	0.002	0.007	0.011
	Female	0.482	0.427	0.007	0.004	0.022	0.002	0.007	0.012
Black	Pooled	0.117	0.021	0.072	0.001	0.008	0.000	0.004	0.009
	Male	0.117	0.026	0.066	0.002	0.010	0.000	0.005	0.008
	Female	0.116	0.017	0.079	0.001	0.006	0.000	0.003	0.010
Asian	Pooled	0.316	0.077	0.006	0.148	0.017	0.001	0.015	0.052
	Male	0.267	0.051	0.003	0.125	0.014	0.001	0.011	0.062
	Female	0.369	0.105	0.008	0.173	0.021	0.001	0.019	0.042
Hispanic	Pooled	0.289	0.083	0.008	0.005	0.144	0.001	0.006	0.043
	Male	0.258	0.073	0.005	0.005	0.131	0.001	0.005	0.038
	Female	0.320	0.093	0.011	0.004	0.156	0.001	0.006	0.048
AIAN	Pooled	0.256	0.132	0.007	0.003	0.020	0.072	0.010	0.012
	Male	0.227	0.119	0.004	0.003	0.017	0.065	0.009	0.010
	Female	0.286	0.146	0.009	0.003	0.023	0.079	0.011	0.014
Other	Pooled	0.299	0.159	0.021	0.021	0.031	0.003	0.043	0.021
	Male	0.269	0.141	0.014	0.022	0.030	0.003	0.039	0.021
	Female	0.328	0.176	0.029	0.020	0.033	0.003	0.046	0.020

Notes: This table presents statistics on rates of interracial marriage separately by sex. For details on construction, see the notes to Table IV.

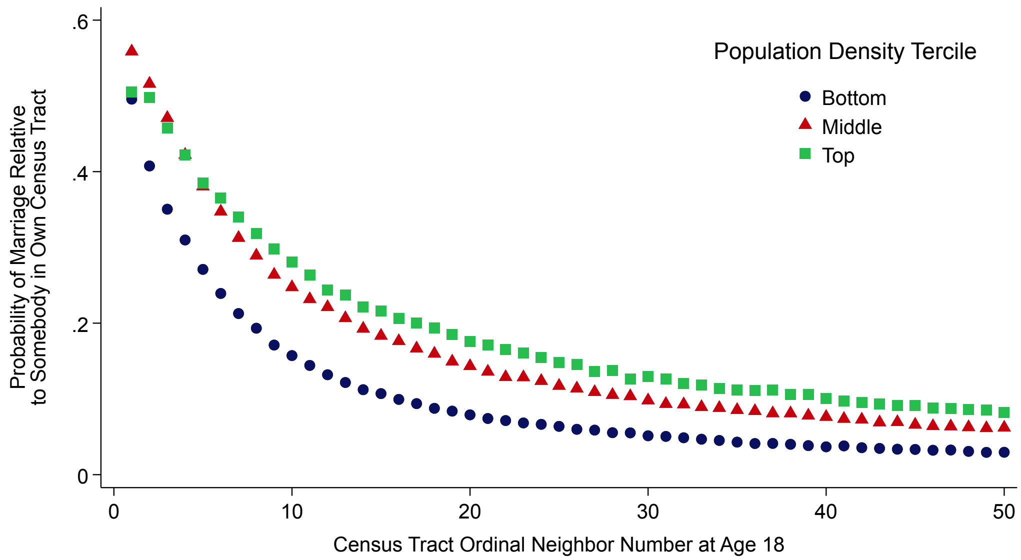
FIGURE A.1: Examples of Nearest 50 Census Tracts



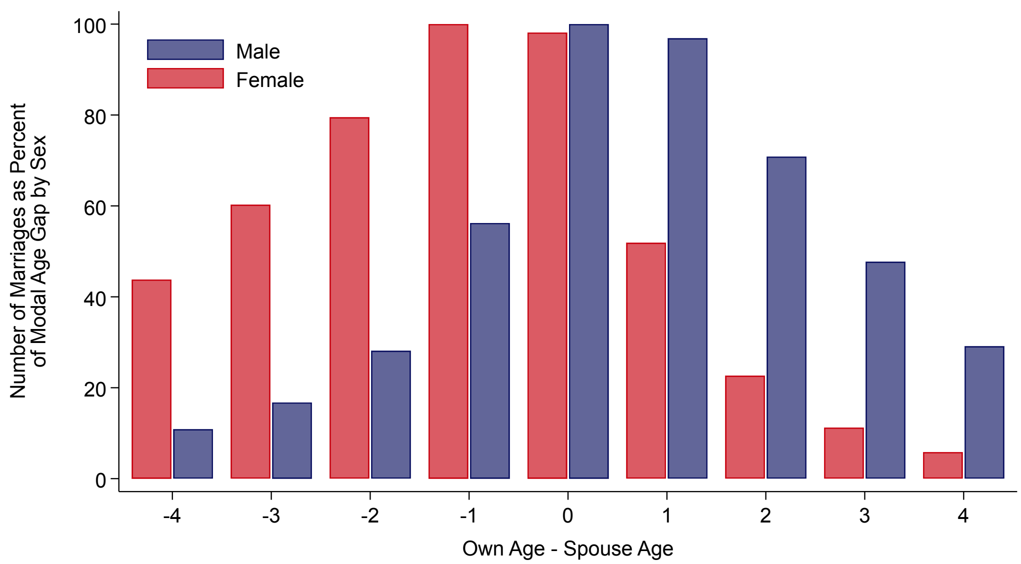
Notes: This figure depicts the nearest 50 Census tract neighbors of a particular focal tract in four U.S. cities. In each case, the neighbors are defined using the distance between Census tract centroids. The focal tract is shown in red, neighbors 1-10 in dark blue, and neighbors 11-50 in light blue. The 2010 FIPS codes for the focal tracts are 25017353700 (Cambridge), 36061005900 (New York), 27137001800 (Duluth), 33001966402 (Gilford).

FIGURE A.2: Distance and Age Weights Used to Construct Exposure Measure

A: Marriage Probability Decay by Distance and Population Density



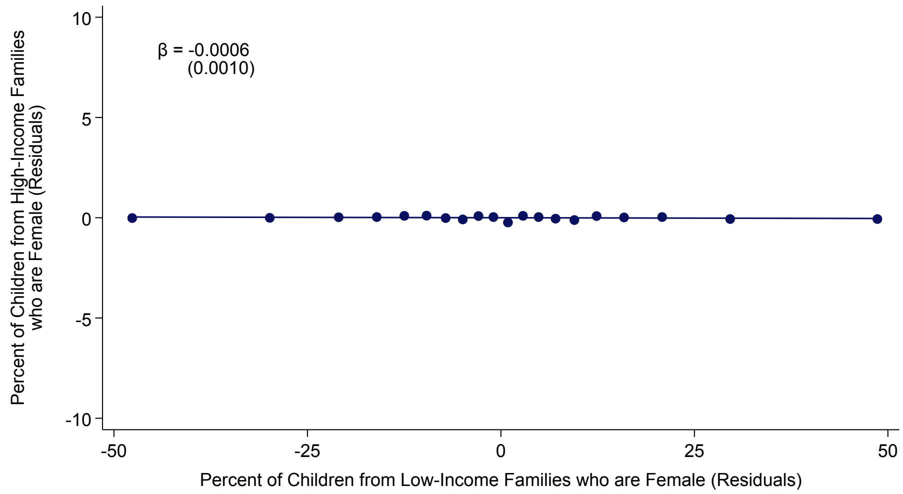
B: Difference in Age From Spouse by Sex



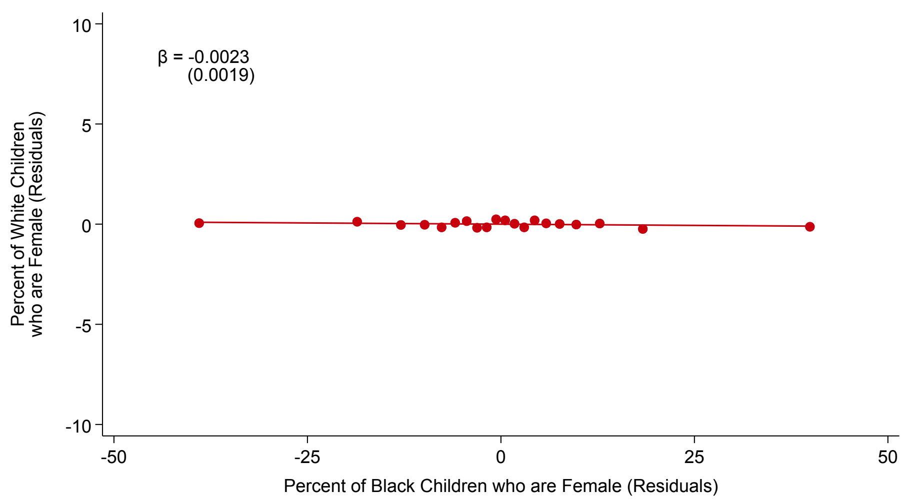
Notes: This figure presents the weights used to construct the exposure measure described in Section III.C. In Panel *A*, we present the distance weights, separately for each tercile of population density, measured in 2000. A pooled version of this figure is shown in Figure II Panel *B*; see figure notes for additional details. Panel *B* presents the frequency of spouse age gaps for married males and females. For each group, the modal age gap is assigned a value of 100. For all other age gaps between -4 and 4, the height of the bar reflects the frequency of the age gap as a percentage of the frequency for the modal age gap.

FIGURE A.3: Within-Neighborhood Correlation of Sex Ratios Across Groups

A: Individuals from Families in the Bottom 25% and Top 25%



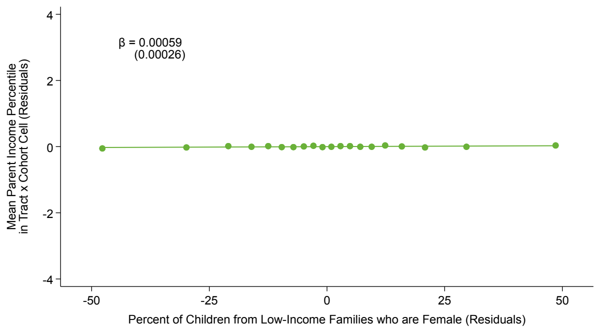
B: White and Black



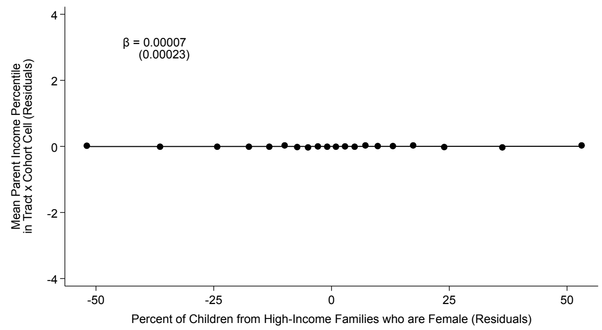
Notes: This figure presents binned scatter plots of the within-neighborhood relationship between the sex-ratios across different class (Panel *A*) and race (Panel *B*) groups. For each Census tract by cohort by income or race group, we calculate the fraction of individuals in the cell who are female. We demean this sex-ratio within Census tract. To generate the plot, we construct 20 bins of the residualized low-income (Black) sex-ratio and plot the mean high-income (white) sex ratio in each bin against the mean low-income (Black) sex-ratio in the bin. The coefficients on the plot are from regressions in a tract by cohort dataset that includes tract fixed effects, with standard errors clustered at the county level.

FIGURE A.4: Tract × Cohort Sex Ratio Balance on Mean Parental Income

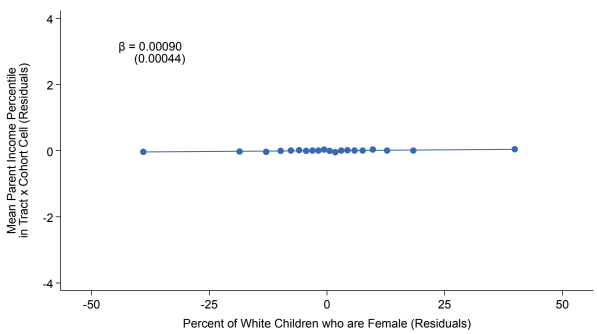
A: Children From Low-Income Families



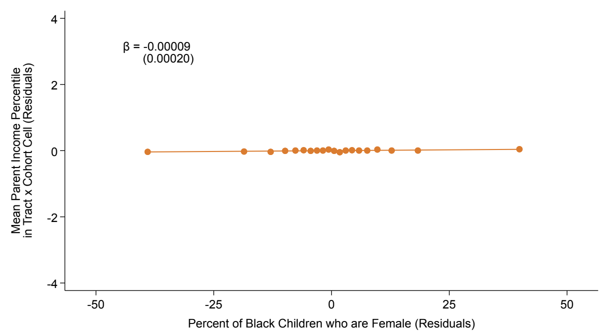
B: Children From High-Income Families



C: White

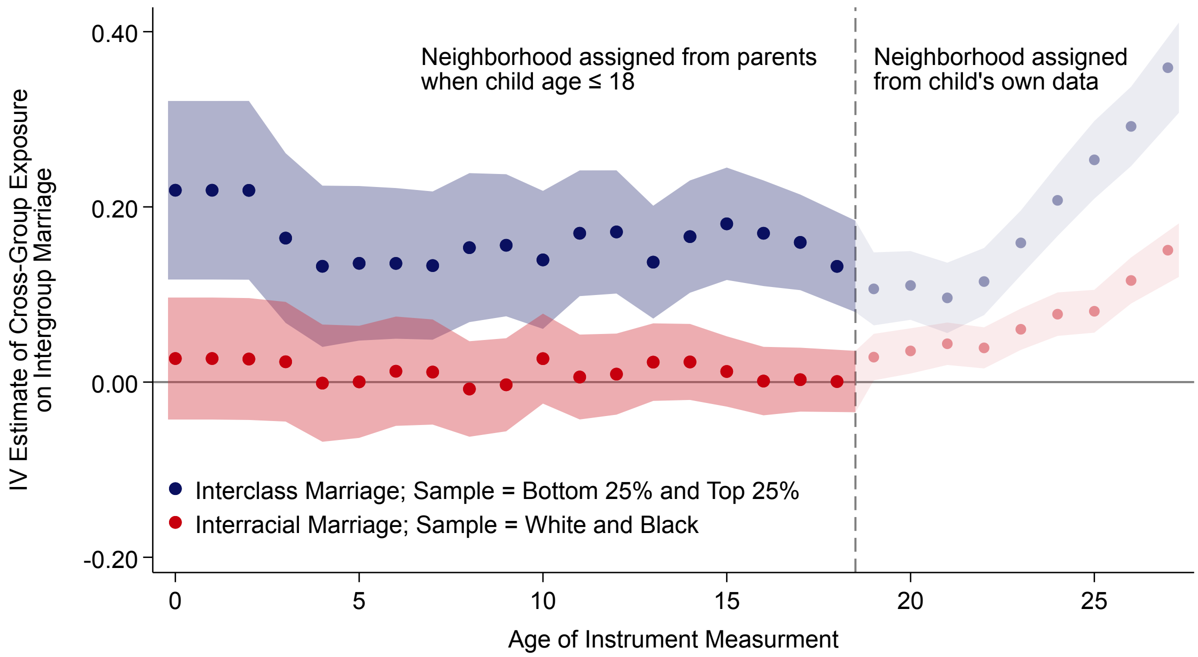


D: Black



Notes: This figure presents binned scatter plots of the within- neighborhood relationship between sex-ratios for specific subgroups and mean parent income rank. We show results for individuals from low-income families (Panel *A*), high income families (Panel *B*), white (Panel *C*), and Black individuals (Panel *D*). In each case, we calculate the fraction of individuals who are female in a Census tract by cohort by family income or race group. We also calculate the mean parent income rank during childhood for that group, using the definition in Section II. To generate the figure, we first construct percentile bins of the number of individuals in the cell. We residualize the sex ratios and the mean parent ranks on Census tract fixed effects and fixed effects for the size bins. To generate the figure, we construct 20 bins of the residualized sex ratios and plot the mean residualized parent income rank against the mean residualized sex ratio in each bin. We report the coefficient from regressions in a tract by cohort dataset of mean parent rank sex ratios controlling for tract and size-bin fixed effects, with standard errors clustered at the county level.

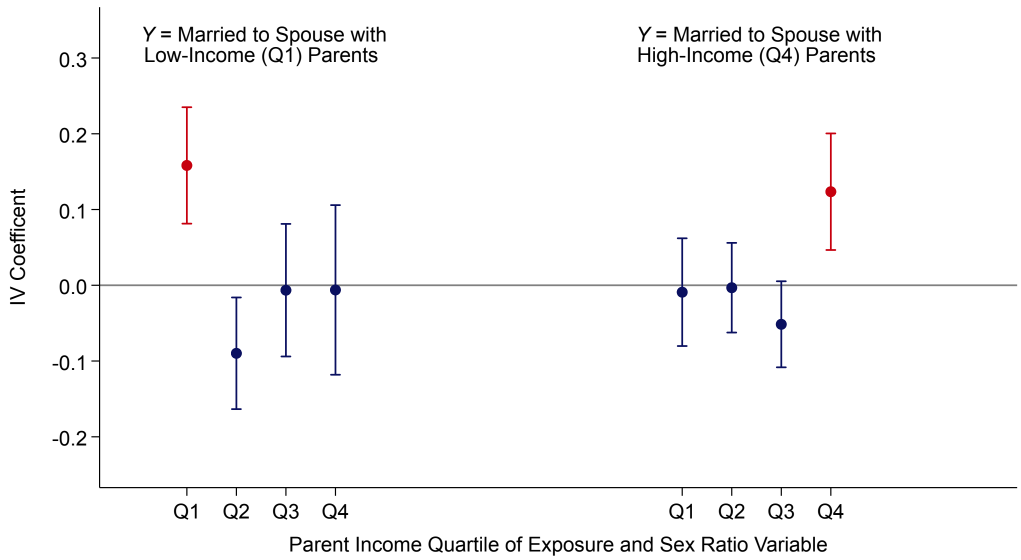
FIGURE A.5: IV Estimates by Age of Instrument Measurement



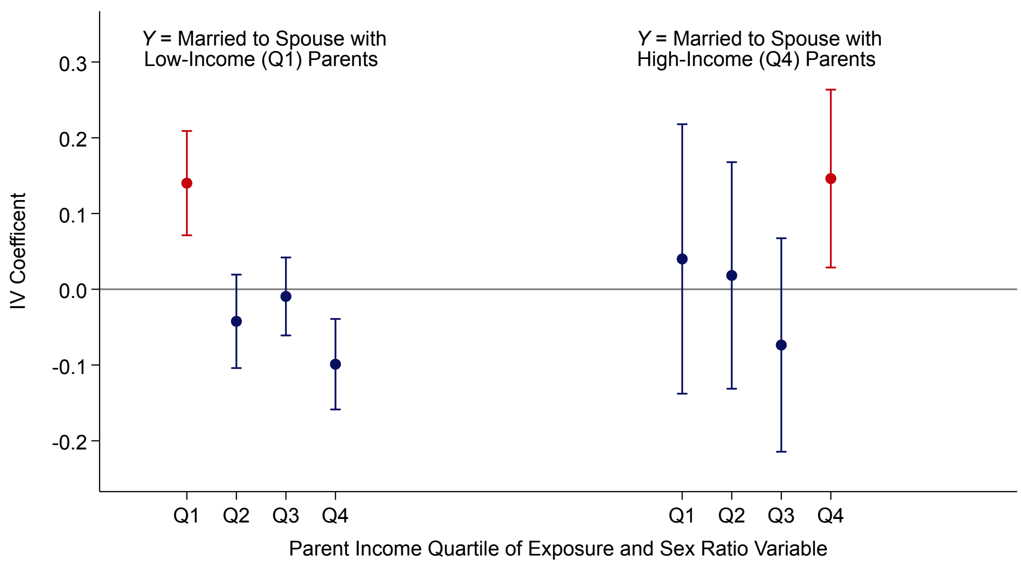
Notes: This figure presents IV coefficients from separate IV regressions where the group-exposure instrument is measured at different ages throughout childhood, separately for class and race. In each case, the outcome is the fraction of individuals with an interclass (interracial) marriage at age 30. In each regression the endogenous exposure is the difference in the fraction of opposite-sex neighbors (measured over ages 18-27) who are from the other group and the fraction of own-sex neighbors who are from the other group. The instrument is the fraction of neighbors in the opposite group who were opposite-sex, where the neighbors are measured at each age of childhood from birth to age 27. Neighborhoods are assigned using parent data through age 18 and child data after age 18. The coefficients where the instrument is measured at age 18 are the baseline results presented in Figure VIII; see Section IV.B for additional details on the estimation procedure.

FIGURE A.6: Impact of Race and Class Exposure on Spouse Characteristics

A: Children with Low-Income (Q1) Parents

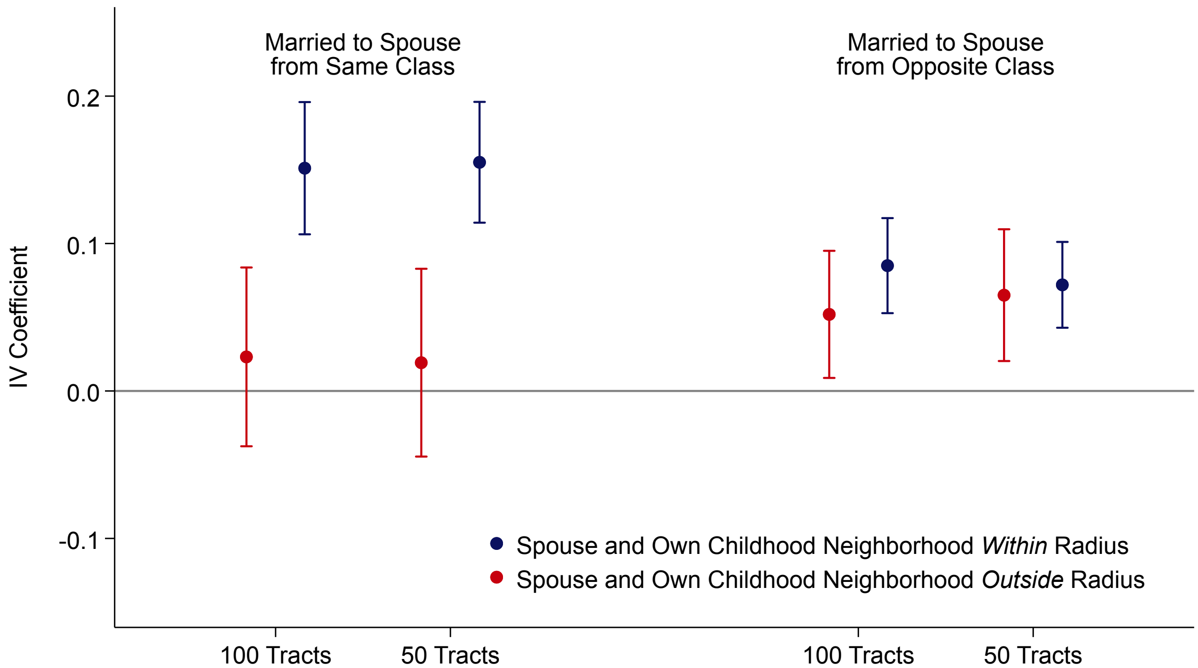


B: Children with High-Income (Q4) Parents



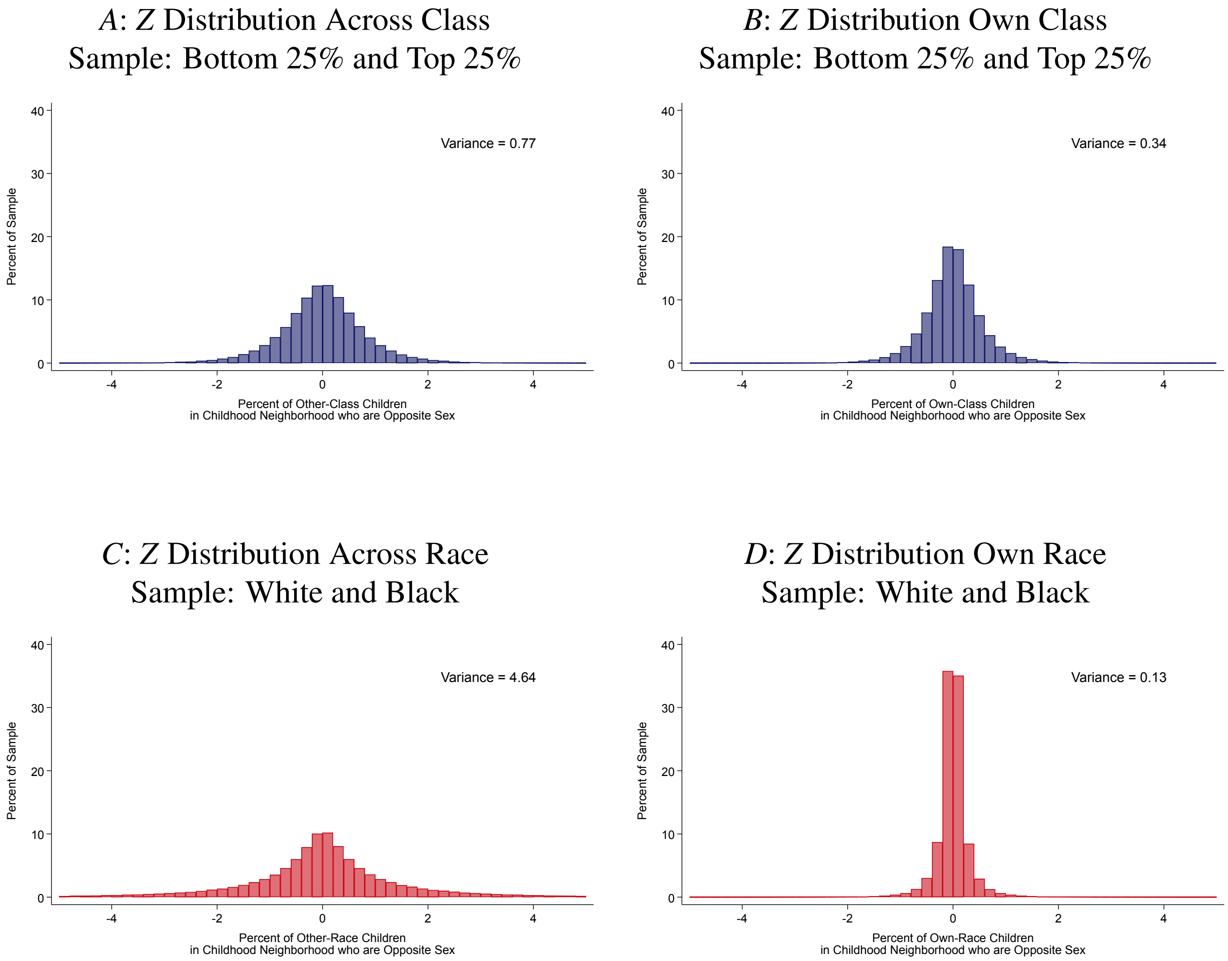
Notes: This figure presents a set of multivariate IV regression for interclass marriage outcomes, separately for individuals from low parent income families (Panel *A*) and high parent income families (Panel *B*). In each case, the marriage outcome, defined above the points, is regressed on four endogenous exposure measures, which we instrument for with four sex ratio instruments, one for each parent income quartile. The controls are the same as in the baseline IV specification; see Section IV.B for additional details.

FIGURE A.7: Within vs. Outside Childhood Neighborhood Marriage
Bottom and Top Quartile



Notes: This figure analyzes whether variation in neighborhood level class-group exposure drives marriage to an individual from the neighborhood vs. outside of the neighborhood. The left half of the figure contains estimates for own-class exposure and having an own-class spouse and the right half uses opposite-class exposure and the effect on having an opposite-class spouse. In each case, the sample contains individuals who come from bottom and top quartile parent income families. In each half of the graph, there are coefficients from four separate regressions. In each case, the right-hand side is the same as in the baseline IV specification; see Section IV.B for additional details on the estimation procedure. The four outcome variables are having a spouse in the class group that is within the 100 nearest neighbor tracts to one's age 18 neighborhood, outside of the nearest 100 tracts, within the 50 nearest neighbor tracts, and outside of the 50 nearest neighbor tracts.

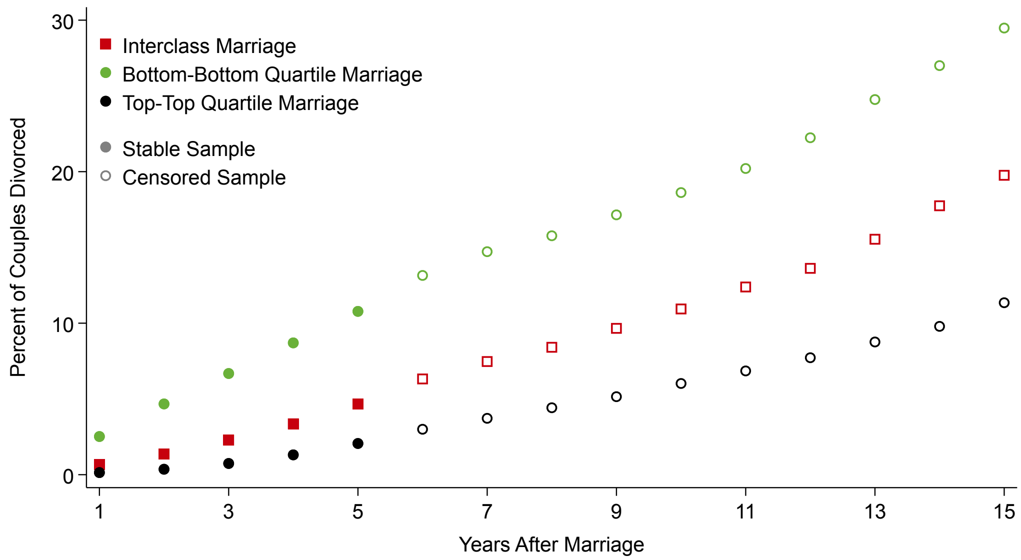
FIGURE A.8: Distribution of Instrument Within vs. Across Groups



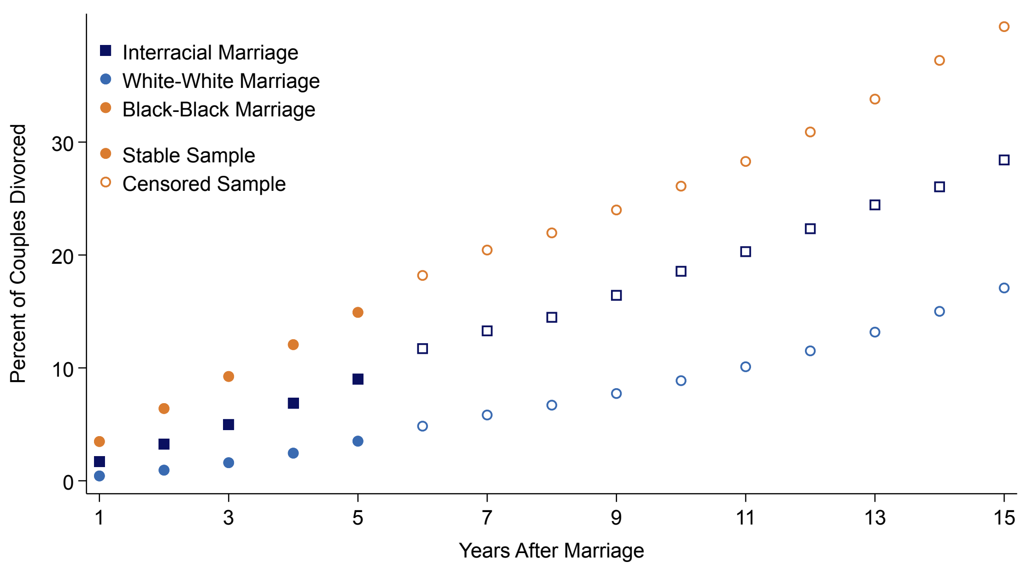
Notes: This figure presents the distribution of four versions of the sex-ratio instrument. In Panels *A* and *B*, the instrument is constructed is using a sample of individuals from the bottom and top 25% of the parent income distribution. In Panel *A*, the instrument is the fraction of neighbors (measured at age 18) from the other class group who are of the opposite sex. In Panel *B*, it is the fraction of own-class neighbors who are the opposite sex. In Panels *C* and *D*, the instrument is constructed is using a sample of white and Black individuals. In Panel *C*, the instrument is the fraction of neighbors from the other race group (e.g. Black neighbors for white indidividuals) who are of the opposite sex. In Panel *D* it is the fraction of own-race neighbors who are the opposite sex. The histograms are analogous to those presented in Figure VII. Panels *A* and *C* are the same histograms to those shown in Figure VII Panels VII*A* and VII*C*. See figure notes for additional details.

FIGURE A.9: Divorce Rates by Marriage Type

A: Interclass vs. Own Class Marriage
Sample: Bottom 25% and Top 25%



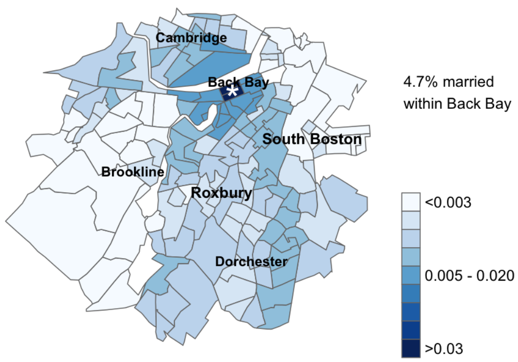
B: Interracial vs. Own Race Marriage
Sample: White and Black



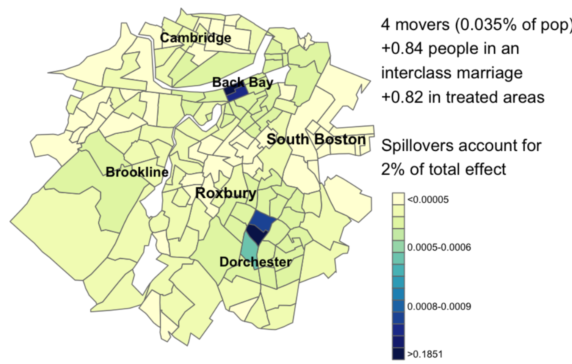
Notes: This figure plots rates of divorce for own-group vs other-group marriage, separately for class (Panel *A*) and race (Panel *B*). The class results are estimated using a sample of individuals from the top and bottom quartile of the parent income distribution. The race results are estimated using white and Black individuals. In each case, the sample is restricted to couples that marry in the year 2013 or earlier. For each year after marriage from 1-15, we plot the fraction of couples that have divorced by that point. Individuals are considered divorced when they file a tax return with a status other than married filing jointly or married filing separately, after previously filing as married. For all couples married in 2013 or earlier, divorce rates are observed 5 years after marriage as the last available year of tax data is 2019. These rates are shown in solid circles. Beyond 5 years, divorce rates are estimated only among couples that married prior to 2013. These censored rates are shown in hollow circles.

FIGURE A.10: Sensitivity of Policy Experiment to Search Costs

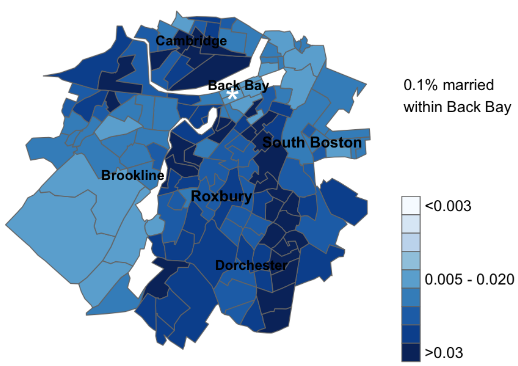
A: High Distance Costs
Percent Married to Resident of Back Bay



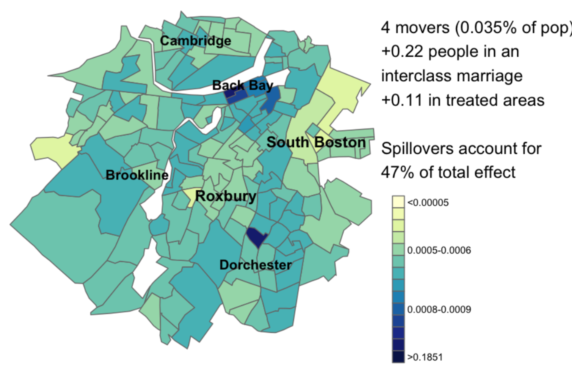
B: High Distance Costs
Δ Interclass Marriage at Neighborhood Level



C: Low Distance Costs
Percent Married to Resident of Back Bay

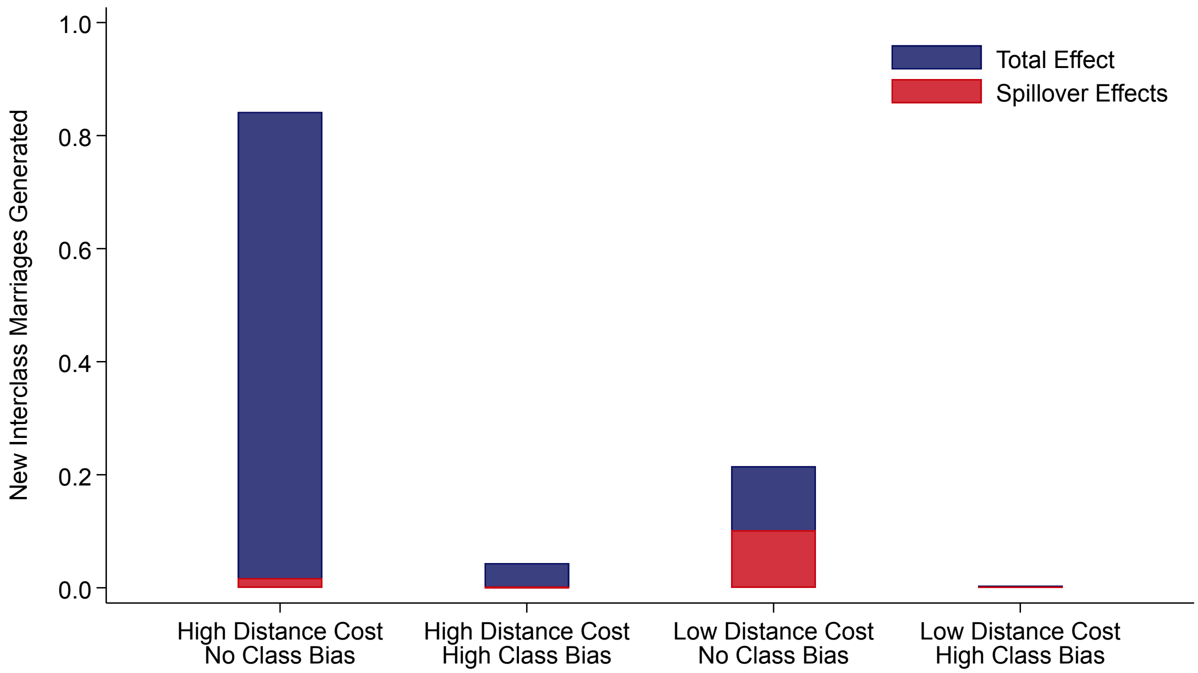


D: Low Distance Costs
Δ Interclass Marriage at Neighborhood Level



Notes: This figure shows how the model responds to changes in the search cost parameter. In Panels *A* and *C* we show the fraction of individuals in each tract married to a resident of Back Bay (denoted by $*$ under high and low search costs, respectively. In Panels *B* and *D*, we show the result of the policy experiment shown in Figure XI Panel *D* and described in Section V.C under high and low search costs, respectively. In each panel, we report the market-wide change in interclass marriage and the change in the two tracts that are directly treated. The difference between these two changes can be explained by spillovers to other neighborhoods. In both the high and low search cost cases, we set class bias to zero for simplicity.

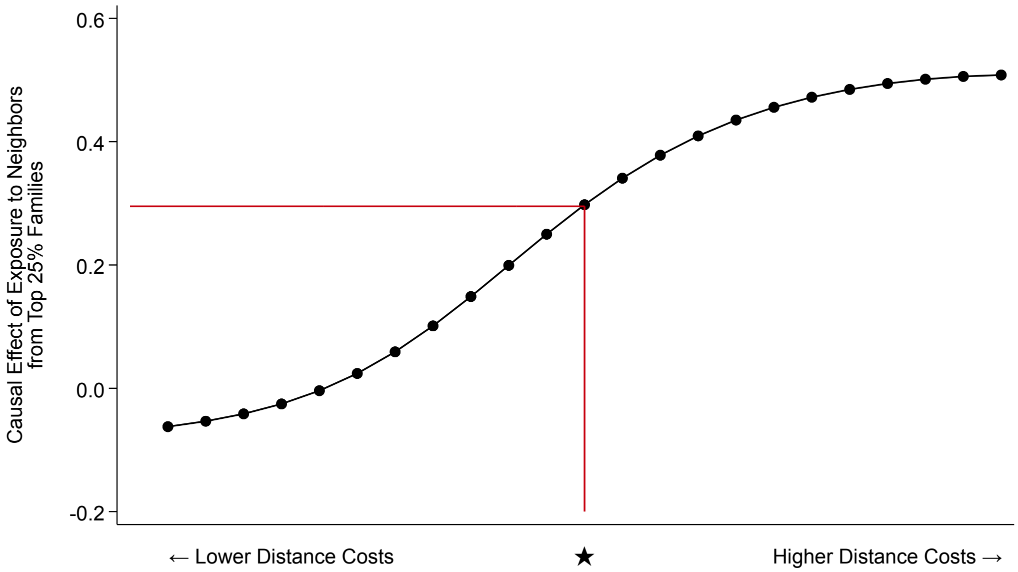
FIGURE A.11: Impact of Policy Simulation as a Function of θ



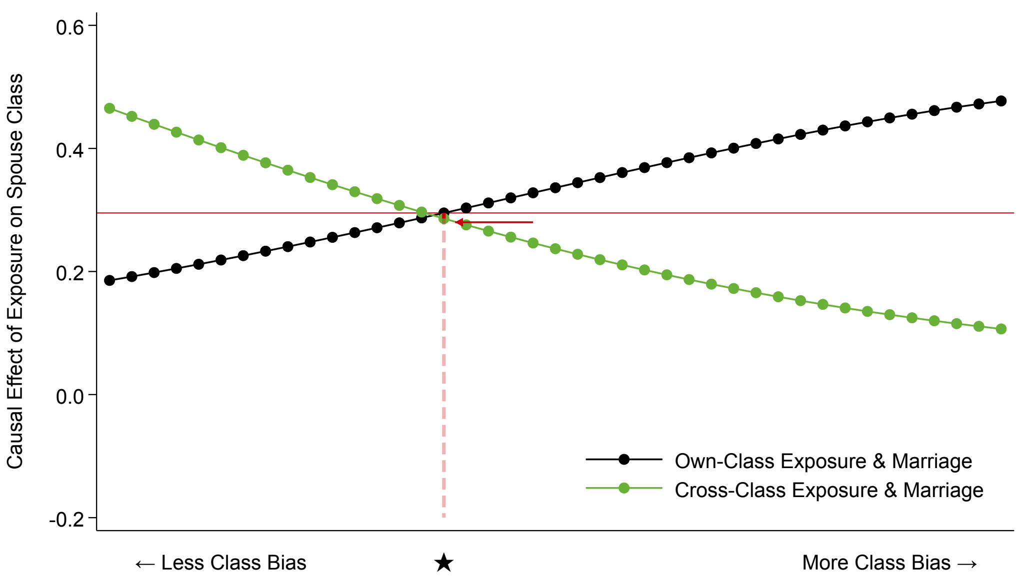
Notes: This figure shows how the results of the policy simulation described in Section V.C and shown in Figure XI Panel *D* are governed by the distance cost and class bias parameters in the model. In each set of bars, we show how the impact of the policy experiment changes interclass marriage in the treated tracts, in blue, and in spillovers to all other tracts, in red. The first and third bars are the same as the simulations shown in Figure A.10 Panels *B* and *D*, respectively.

FIGURE A.12: Relationship Between Model Parameters and Simulated Partial Equilibrium Moments

A: Own-Class Exposure Effect to Identify Search Costs



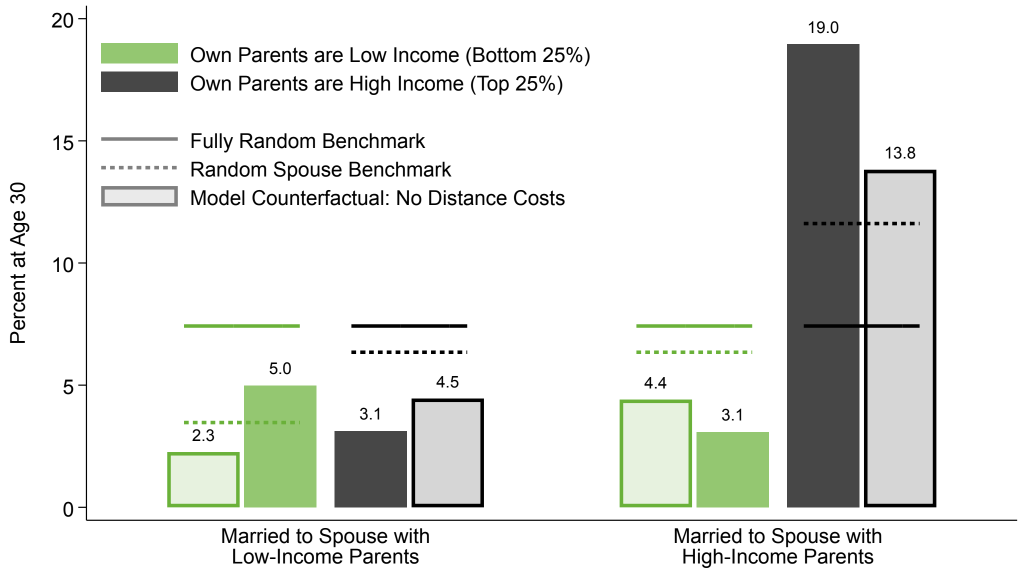
B: Other-Class Exposure Effect to Identify Class Bias



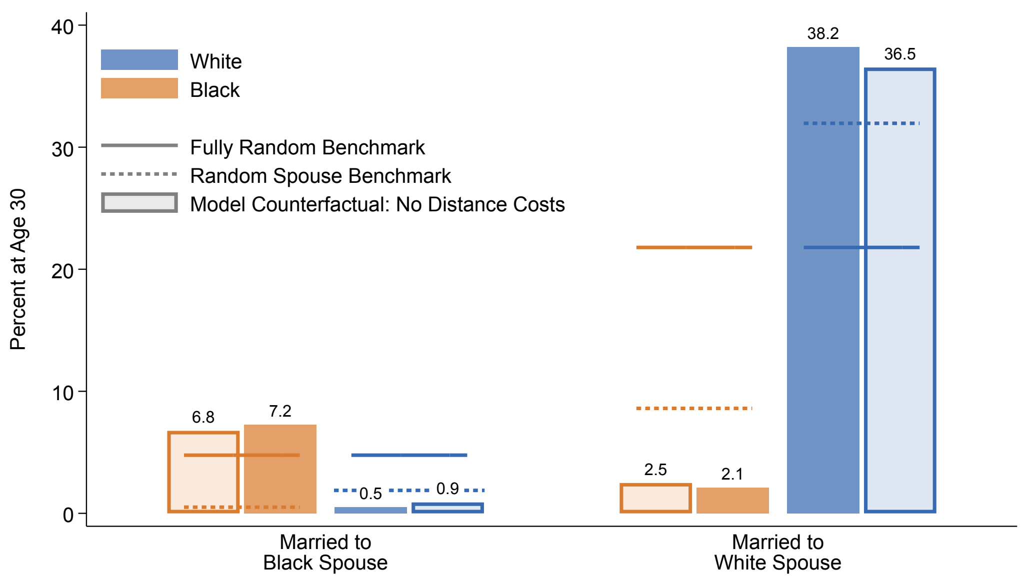
Notes: This figure highlights the logic used to identify the distance cost and class bias parameters in the spirit of Andrews et al. (2017). In Panel *A* we show how the estimates for the effect of own-class exposure change in the model as the search cost varies. The true moment value, shown via the horizontal line, pins down the search cost value at ⋆. In Panel *B*, we show how the own- and other-class estimates change as a function of the class bias parameter. We set the intercept of the own-class series such that it crosses the true value of the estimate at the point where the true value of the difference between own and other class estimates also matches the data, as shown via the arrow. Setting the intercept in this way pins down the level class bias at ⋆ as shown.

FIGURE A.13: Impact of Search Costs on Marital Homophily by Class and Race

A: Rates of Interclass vs. Within-Class Marriage



B: Rates of Interracial vs. Within-Race Marriage



Notes: This figure summarizes key patterns of marital homophily by class (parent income) and race as well as counterfactuals from a spatial model of the marriage market. The solid bars and the two benchmarks shown are repeated from Figure I; see figure notes for additional details. The hollow bars present a counterfactual from the model that removes any search costs across neighborhoods. In the baseline model, individuals have preferences to marry a nearby neighbor, motivated by Figure II, that decay over the nearest 50 Census tracts. In the counterfactual, those preferences are set to zero, removing the effect of residential segregation on the pool of desirable spouses. Importantly, the group specific marriage rates (e.g. the overall fraction of individuals from low parent income families who are married) are held constant in the counterfactual. For additional details on the model setup and the counterfactual, see Section V.D.