The Melting Pot: Industrialization and Integration in America

Arkadev Ghosh, Sam Hwang, Munir Squires*

March 21, 2025

Abstract

Industrialization happened alongside a fundamental reshaping of American society, as closed, localized communities transformed into more open, interconnected networks. This paper demonstrates that industrial expansion played a causal role in increasing social integration, creating a "melting pot" effect where traditional community boundaries weakened. To measure this social integration, we introduce *marital distance*, a novel metric that captures the network distance separating spouses. Intuitively, this measure captures whether marriages reinforced existing social ties (shorter distances) or bridged previously distant social circles (longer distances). Using genealogical data covering 50 million Americans (1750–1950), we document that social integration gradually increased, with marital distance rising from an average of 10 steps in 1750 to 16 steps by 1950, with the sharpest increase coinciding with America's period of intensive industrialization (1870-1915). A shift-share analysis shows that manufacturing growth drove integration both by mixing individuals from varied communities through migration and by reducing social barriers within locations. Our highly connected dataset, where approximately 95% of married couples can be linked through some pre-existing path, provides unprecedented insight into how economic modernization created a more integrated society by expanding primary bonds beyond traditional community boundaries.

1 Introduction

America's transformation from a collection of distinct communities into a 'melting pot' is a defining feature of its national development (Abramitzky et al., 2014; Mohnenab and Viarengocde, 2023). This process of social integration—the formation of meaningful connections across traditional boundaries—fundamentally reshaped American society during industrialization. While the melting pot metaphor has long captured public imagination (Zangwill,

*The authors are from Duke University (Ghosh) and University of British Columbia (Hwang and Squires). This research was supported by the Social Sciences and Humanities Research Council of Canada (SSHRC). The authors declare that they have no financial relationships or other potential conflicts of interest related to this work.

1921), measuring this integration systematically has remained challenging, especially in historical contexts.

Economic modernization weakens traditional community boundaries and facilitates more open, interconnected networks (Weber, 1951; Enke, 2019; Polanyi, 1944). As industrial capitalism replaced traditional agrarian economies, Americans increasingly formed connections that bridged previously separate social circles. Yet systematic empirical evidence on this process has remained scarce due to the difficulty of measuring historical social networks at scale.

This paper establishes a causal relationship between industrialization and social integration by introducing a novel measure of social connectedness: *marital distance*. This network-based metric quantifies the number of steps separating two spouses in a social network. Specifically, we construct a network connecting individuals through family ties, then measure the shortest path between spouses before their marriage. This captures whether marriages reinforced existing social ties (shorter distances) or bridged distant parts of society (longer distances). Intuitively, everyone in a tight-knit community is connected through relatively short network paths — not because they are blood relatives, but because of previous marriages connecting their extended networks. Marrying within such a community means implies a short marital distance. Conversely, marriages crossing between separate social clusters exhibit longer distances, indicating connections that span rather than reinforce existing community boundaries. Higher marital distances signal a more integrated society where marriages connect distant parts of the social network instead of remaining within closed circles.

Marriage is particularly well-suited for measuring social integration. It represents the most significant chosen relationship in a person's life, reflecting deeply held social preferences while requiring substantial commitment. Crucially for our purposes, it is also the only chosen relationship systematically recorded across the population in historical data. Marriage connects entire family networks, potentially bridging previously separate social circles, and these new family units shape future generations.[1]

Our measure offers exceptional historical coverage and granularity. Using genealogical data on marriages from 1750 to 1950, we analyze over 270,000 county-year observations. Crucially, marital distance also captures individual-level variation. By linking genealogical records to US census data, we study how marital distance correlates with economic outcomes, occupational transitions, and geographic mobility.

To distinguish between changes in network structure from changes in social preferences, we construct counterfactuals representing what we think of as maximal integration. By measuring the marital distance of randomly matched couples, we establish benchmarks where social networks are ignored when making marriage choices. The gap between actual and random distance measures the intensity of social barriers. By constructing both within-county and national random benchmarks, we can separate geographic sorting effects (people living in homogeneous communities) from within-location preferences (selecting socially proximate partners within diverse communities).

[1]This focus aligns with Gordon (1964)'s seminal work on assimilation in America, which argues that while cultural integration may occur for most immigrant groups, full structural integration happens through intermarriage.

However the reliability of any network-based measure depends on the quality of the underlying data. Our marital distance measure draws on a genealogical network that is exceptionally interconnected: 86% of individuals belong to a single vast family tree rather than isolated smaller ones. More importantly, 95% of married couples we analyze can be linked through pre-existing family ties. This connectivity allows us to measure network distance between spouses.

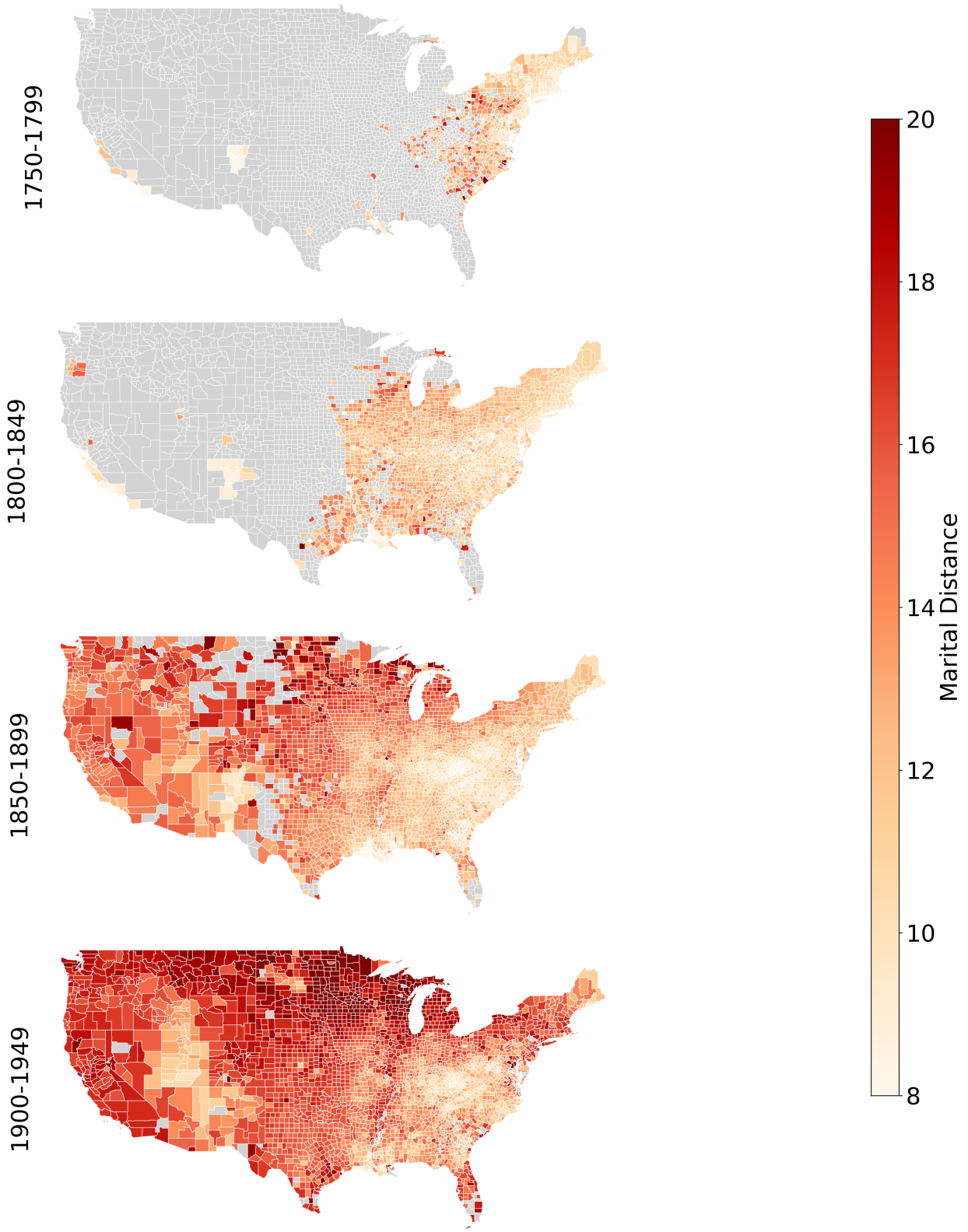
For marital distance to serve as a meaningful proxy for social integration, it should correlate with established metrics of social cohesion. Indeed, we find for example that counties with higher marital distances exhibit lower Local Name Index scores, indicating reduced in-group identification through distinctive naming patterns Raz (2025); Fryer and Levitt (2004). The relationship is stronger with Intra-Community Marriage rates, which measure the share of couples with common birthplaces, a phenomenon Triandis (1995) established as characteristic of close-knit communities. These correlations remain robust when controlling for county and time fixed effects. Our measure also correlates negatively with cultural tightness indicators Gelfand et al. (2006) and metrics of kinship density Nelson (2020). Counties with shorter historical marital distances in the 1940s exhibit higher social network density in contemporary Facebook data collected nearly 80 years later Chetty et al. (2022), showing temporal stability in the social structures our measure captures.

Finally, a robust measure should predict relevant outcomes beyond its immediate domain. We show that those who marry at greater social distances have increased post-marital geographic mobility, are less likely to live on farms, and have higher occupational income scores. These patterns persist when comparing siblings from the same family. The within-family approach helps address concerns that our measure might simply capture family-level selection, showing that marital distance identifies social connectivity that shapes economic trajectories independent of shared background. By comparing siblings from the same household who marry at different network distances, we isolate the relationship between marital distance and economic outcomes from confounding factors like family resources, cultural norms, and geographic constraints.

While marital distance provides unique insights into historical social integration, it has several limitations. First, our data only captures Americans with extensive genealogical records. Notably, non-whites are underrepresented, limiting our ability to analyze integration across racial lines (Hwang and Squires, 2024; Price et al., 2021). Linking our genealogical records to the census allows us to document which individuals are missing, but cannot recover network connections for these underrepresented groups. Second, couples with no pre-existing connections in our genealogical network pose measurement challenges. We present multiple methods to deal with these cases, which represent about 5% of marriages in our sample. Third, like all network-based measures, marital distance is sensitive to missing data. To mitigate this concern, we restrict analysis to individuals with genealogical links to all four grandparents. Despite these limitations, we argue that marital distance offers a powerful lens for studying social integration in the US across a wide timespan.

We document a consistent tendency toward socially proximate marriages throughout American history, with individuals marrying spouses closer in the network than expected under random matching—even when limited to the same county. Between 1750 and 1950, two important trends emerged. First, average marital distance increased from about 10 to 16 steps, with the sharpest rise during America's intensive industrialization (1870-1915).

Figure 1: Marital distance (social integration) by county, 1750–1949.



Note: This figure shows the average marital distance by county from 1750 to 1949, in 50-year periods. Darker shades indicate higher marital distances (more integration), while lighter shades represent lower values (more clustering).

Second, the gap between actual and random marriage distances gradually narrowed over this period. Together, these patterns provide strong evidence of increasing social integration during economic transformation. Regional variation was substantial, as illustrated in Figure 1, with higher marital distances in the northern and western United States compared to the South and Appalachia, reflecting how economic development and cultural factors shaped marriage networks across different regions.

We demonstrate that industrialization drove this integration through county-level panel analysis from 1850 to 1940. Counties with higher manufacturing employment experienced greater increases in marital distance, even after accounting for urbanization, transportation infrastructure, and demographic shifts. To address endogeneity concerns that manufacturing location decisions may respond to local social conditions, we employ a shift-share instrumental variables strategy based on initial industrial composition and national trends (Bartik, 1991; Goldsmith-Pinkham et al., 2020). The IV estimates exceed OLS results, suggesting standard approaches understate manufacturing's causal effect on social integration. A ten percentage point increase in manufacturing employment raises marital distance by approximately 1.2 steps, with stronger effects in previously insular communities and in industries employing women.

How much of this effect reflects changing marriage preferences versus shifts in the available pool of potential spouses? To disentangle these mechanisms, we include random marital distance as a control in our regressions, effectively measuring how the network structure itself constrains marriage options independently of preference changes. We find that one-quarter of the manufacturing effect stems from changes in spouse selection preferences within existing populations, while three-quarters results from compositional changes in the marriage pool through migration. Manufacturing increased marital distance primarily by fragmenting family networks: migrants arrived as nuclear families or individuals, separated from their extended kin. Newcomers married at greater social distances simply because close kin were no longer locally available. Manufacturing growth facilitated social networks that spanned traditional community boundaries both by changing the pool of local marriage partners and by shifting preferences toward more socially distant marriages.

Our findings contribute to three strands of literature. First, we provide direct empirical evidence on how family networks evolved during industrialization, advancing research on social structure and economic development (Alesina and Giuliano, 2014; Greif, 2006; Enke, 2019). Second, we offer a new perspective on intermarriage and social assimilation by measuring social distance using realized family networks rather than predetermined categories (Gordon, 1964; Kalmijn, 1998). Third, we connect the study of social connectivity to research on economic mobility (Chetty et al., 2014), showing how the expansion of social networks facilitated movement into industrial labor markets by weakening geographic constraints on economic opportunity.

By systematically measuring social connectedness over two centuries of economic transformation, this paper shows how industrialization not only changed what Americans produced but fundamentally altered who they married, where they lived, and how they formed connections across previously separate social circles. These findings provide new empirical support for the hypothesis that economic modernization reshapes social relationships, facilitating greater social integration by expanding networks beyond tight-knit community structures toward more open, interconnected social systems that became characteristic of American

society.

2 Data and Network Construction

2.1 Genealogical Data

Our analysis draws on a novel dataset of genealogical records from FamilySearch, a platform where users contribute and validate historical family information (Price et al., 2021; Hwang and Squires, 2024). Our dataset contains detailed records of births, deaths, marriages, and family relationships for approximately 50.9 million individuals in the United States from 1750 to 1950. A key feature of these data is that they allow us to observe an interconnected family network where most individual family trees link together into a vast genealogical structure.

The network we construct comprises 96.3 million links, including 13.3 million marriage relationships and 83 million parent-child relationships (41.5 million father-child and 41.5 million mother-child). As shown in Table 1, approximately 85.8% of individuals belong to one large connected component rather than isolated family trees, with a maximum network distance of 82 steps. This high degree of connectivity is crucial for our analysis as it enables us to measure social distances between most potential marriage partners.

Table 1: Network Size and Structure

Network Overview	Count
Number of individuals (nodes)	50.9M
Number of links (edges)	96.3M
Marriage	13.3M
Father-child	41.5M
Mother-child	41.5M
Main Connected Family Tree	
Size of largest component (nodes)	43.7M
Percentage in largest tree	85.8%
Maximum distance	82
Known Grandparents per Individual	
0 grandparents	30.0%
1-3 grandparents	17.9%
4 grandparents	52.1%

A notable feature of our data is that over half (52.1%) of individuals have all four grandparents identified. As is common when working with genealogies, we focus on a subset of our data with sufficiently complete family trees (Blanc, 2024). Specifically, we restrict attention to individuals with complete grandparent information to reduce the likelihood that our marital distance measures are artificially truncated by missing data. This restriction helps mitigate the possibility that apparently distant connections are simply a result of incomplete genealogical records.

2.1.1 Data Quality and Coverage

To assess the quality and representativeness of our data, we link genealogical profiles to census records. This allows us to compare characteristics of individuals in our sample to the broader population. While our data cover a substantial portion of the U.S. population, roughly 60% of individuals born before 1900, the coverage is not uniform across groups.

The most significant limitation is that non-whites are severely underrepresented, comprising only about 1% of our sample compared to 10% of the census population (see Hwang and Squires (2024) for more details). This reflects both limited historical record-keeping for non-white populations, particularly before emancipation, and differential inclusion in family history research. Given this limitation, our results should be interpreted as primarily reflecting patterns among white Americans.

Our sample also overrepresents native-born individuals and farmers relative to the census population. These differences in coverage inform how we construct our analysis sample and interpret our results. To address potential bias from selective coverage, we implement sample restrictions and robustness checks in our main analyses, such as focusing on counties with high and consistent data coverage over time.

Finally, missing network observations can alter fundamental structural properties, introducing unpredictable biases in connectivity estimates. We evaluate the consequences of incomplete genealogical information by comparing network statistics such as average degree and connectivity against external benchmarks derived from census records. This analysis examines whether individuals with partial genealogical links exhibit systematically different characteristics from those with more complete family trees, which could distort our inferences about marriage patterns and social structure. This approach enables us to quantify potential measurement errors arising from differential data availability across regions and time periods, ensuring that our measures of social distance and connectivity remain methodologically sound.

2.2 Linkage to Census Records

To examine the relationship between marital distance and economic outcomes, we link individuals in our genealogical dataset to census records from 1850 to 1940. This linkage allows us to observe occupation, place of residence, household composition, and other socioeconomic characteristics. The linked dataset enables us to track how individuals with different marital distances fared in terms of geographic mobility, occupational transitions, and economic advancement.

For our individual-level analyses, our main sample includes approximately 1.8 million individuals with post-marriage census records. Many individuals appear multiple times in our analysis as they are observed across different census years, allowing us to track outcomes over the life course.

A strength of our genealogical data is that it makes it straightforward to identify siblings (Hwang et al., 2024). Specifically, we compare siblings who married at different network distances. While imperfect, this approach helps isolate the effect of marital distance from broader family background influences, addressing potential concerns about selection into distant marriages.

2.3 Manufacturing Data

Our data on manufacturing employment come from county-level records compiled by Haines and Inter-university Consortium for Political and Social Research (2010), which provides information on industrial development from 1850 to 1940. We digitized these records to create a panel of manufacturing employment at the county level. Our main measure is the share of the county population employed in manufacturing.

Coverage of manufacturing data improves over time, with more detailed industry breakdowns available in later years. We use this industry-level detail to construct our shift-share instrument, which we discuss in more detail in Section 5. For some analyses, we supplement the county-level data with information about specific manufacturing establishments from the U.S. Census of Manufactures as discussed in Hornbeck et al. (2024), though these more detailed data are only available for the 1850, 1870, and 1880 census rounds.

3 Measuring Marital Distance

3.1 Constructing the Family Network

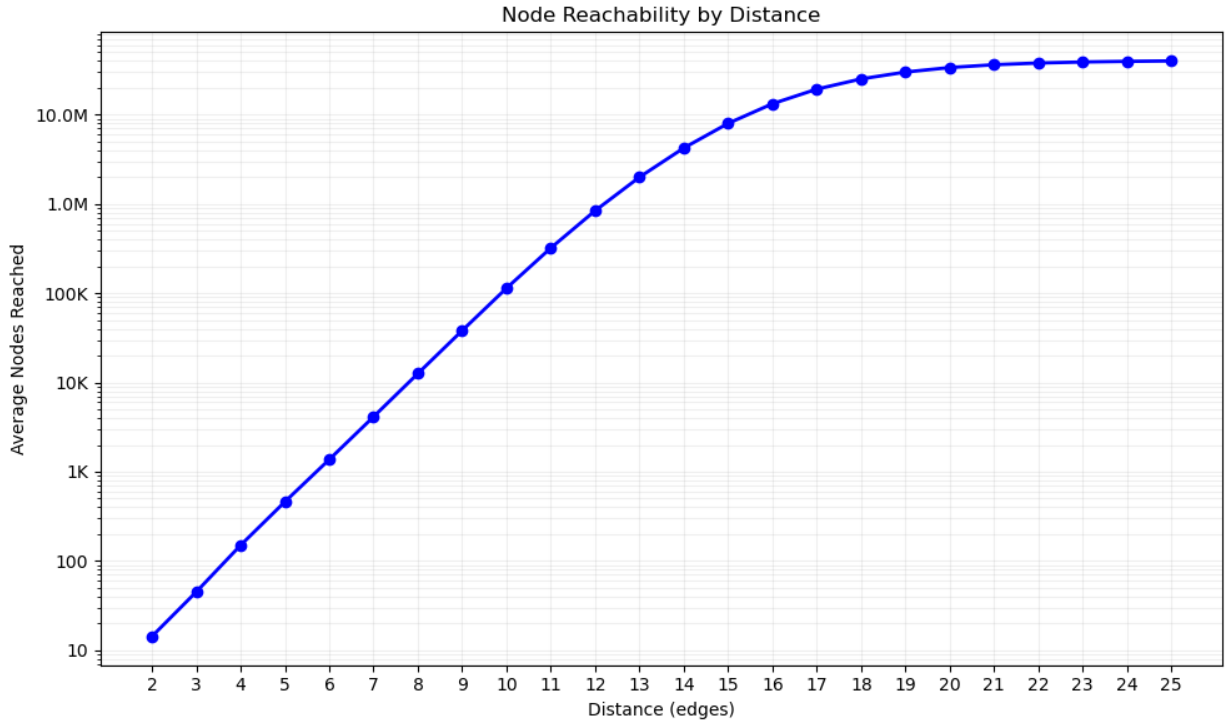
To measure social integration through marriage patterns, we construct a comprehensive historical family network using genealogical data spanning from 1750 to 1950. The network represents individuals as nodes connected by two fundamental types of relationships: parent-child ties and marriages. Parent-child connections form the vertical structure of family trees, while marriages create horizontal bridges between lineages. This allows us to trace both blood relations and marriage-based connections across multiple generations. The resulting network includes approximately 50 million individuals connected through over 13 million marriage relationships and more than 83 million parent-child relationships.

Figure 2 illustrates the expansion of reachability within the family network as a function of distance. To construct this measure, we randomly select nodes and count how many other nodes they can reach by traveling along network edges at increasing distances, repeating this process 1,000 times and averaging the results. The reachability pattern follows an exponential trajectory: roughly speaking, each additional two steps of distance increases the number of reachable nodes by a factor of ten. While a node can typically reach about 50 relatives within three steps, by five steps this increases to approximately 500 nodes, and by 10 steps, this number exceeds 100,000. This exponential growth begins to attenuate around distance 13, and the number of reachable nodes effectively plateaus around distance 22, by which point most nodes in the connected component of the network become accessible.

3.2 Defining Marital Distance

For each marriage, we define *marital distance* as the length of the shortest path connecting the spouses through the family network that existed prior to their marriage. Figure 3 illustrates this concept using a stylized family network. In this example, the shortest path connecting the husband and wife (shown in red) passes through seven steps. Starting from the husband, the path goes through his mother (Mo), grandmother (MoMo), uncle (MoBr), the uncle's wife (wife's FaSi), her father (FaFa), his son (Fa) and finally his daughter, the

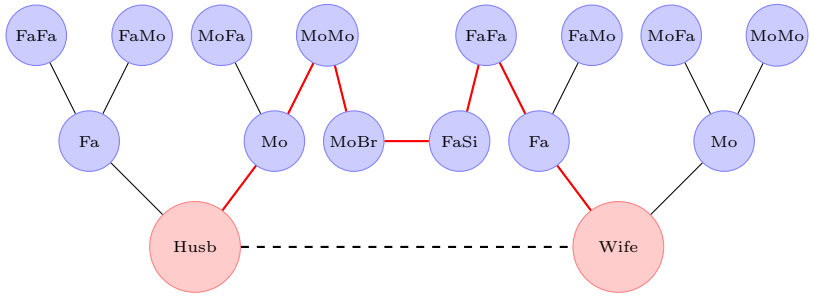
Figure 2: Node reachability by marital distance in the family network



Note: This plot shows the average number of individuals reachable within a given marital distance in the family network, on a logarithmic scale.

wife. Each step represents either a parent-child relationship or a marriage that existed prior to the couple's own marriage (shown as a dashed line).

Figure 3: Stylized shortest path connecting spouses



This example demonstrates how marital distance can capture various types of family relationships. For comparison, first cousins would have a distance of 4 (connecting through shared grandparents), second cousins a distance of 6 (connecting through great-grandparents), and third cousins a distance of 8. The path in our example, with length 7, indicates a relationship more distant than first cousins but closer than third cousins, though following a more complex route through both blood relations and in-law ties rather than solely through common ancestors.

The example also shows how marital distance captures social connections beyond simple consanguinity. Rather than only measuring blood relationships, our measure incorporates the full range of family ties, including connections formed through previous marriages.

Importantly, we do not assume spouses are aware of the specific paths that connect them in the family network, particularly for more distant connections. In fact, when examining the actual paths connecting couples in our data (as shown in Appendix Figures A.1 to A.6), many connections span numerous steps through complex family relationships that would be virtually impossible for individuals to trace. Rather, marital distance serves as an objective measure of social proximity: individuals who are closer in the network are more likely to share social contexts, cultural values, and mutual acquaintances, even if they cannot articulate the exact genealogical ties between them. The value of marital distance lies in capturing these structural dimensions of social integration, regardless of whether the individuals involved consciously perceive these connections.

Computing marital distance requires constructing a historical subgraph containing only relationships formed before the marriage year. The shortest path algorithm explores all possible connections between spouses through two types of pre-existing relationships: blood relations (paths through shared ancestors, such as cousins connected via grandparents) and in-law relationships (paths through previous marriages that created family bridges). We treat these relationships as undirected.

This temporal restriction ensures that we measure authentic social distances at the time of marriage, preventing back-projection of future relationships. For instance, when measuring the distance between spouses married in 1850, we only consider family connections that existed before their marriage, excluding any children born to the couple or subsequent marriages of their relatives. Each relationship in our network has a temporal attribute: parent-child edges are tagged with the child's birth year, while marriage edges are annotated

with the marriage year when available, or the birth year of the first child as a proxy when the marriage date is unknown.

To ensure measurement quality, we impose additional restrictions. Most importantly, we require that both spouses have all four grandparents identified in our data before calculating their marital distance. This restriction helps ensure that apparently distant connections are not simply artifacts of missing genealogical information.

Marital distance effectively quantifies whether marriages primarily reinforced existing kinship structures (short distances) or bridged previously distant families (long distances). The systematic variation in these distances across time and space provides insights into evolving patterns of social integration through marriage.

3.3 Random Marital Distance

To interpret observed marital distances, we need a benchmark for comparison. We construct this benchmark by generating counterfactual marriages that represent what marital distances would be if individuals selected spouses without regard to their position in the social network. The comparison between actual and random marital distances reveals the extent to which marriage choices reflect social structure beyond what would occur by chance.

We implement two complementary versions of random matching, carefully preserving key demographic constraints while relaxing social network constraints. In all random pairings, we only pair males with females who were active in the marriage market in the same year, and we only consider individuals with complete genealogical information (all four grandparents are in our data).

3.3.1 Within-County Random Matching

Our first approach constructs within-county random matches, pairing individuals who married in the same county-year. For each county in a given year, we identify all males and females who actually married that year, creating pools of potential husbands and wives. We then randomly generate alternative pairings from these pools, excluding actual marriages that occurred. This approach accounts for geographic constraints and local marriage markets, capturing the extent of social sorting conditional on location. It maintains the county-level demographic composition of the marriage market while randomizing connections within those boundaries.

Formally, for each county c and year t, we define sets M_{ct} and F_{ct} as all males and females, respectively, who married in that county-year. We then construct the set of potential random marriages as the Cartesian product $M_{ct} \times F_{ct}$, excluding the set of actual marriages A_{ct}. From this set of potential matches, we randomly sample up to 50 pairs per county-year to create our counterfactual marriage dataset.

This approach preserves local demographic constraints while removing social network preferences in spouse selection. By comparing actual marriages to within-county random marriages, we can isolate the effect of social network position on mate selection, controlling for the geographic constraints of local marriage markets.

3.3.2 National Random Matching

Our second approach creates counterfactual marriages from the national pool of individuals who married in the same year, regardless of location. For each year, we identify all males and females who married that year, creating national pools of potential spouses. We then randomly pair males and females from these pools, generating marriages that could have occurred if geographic and social constraints were removed.

Formally, for each year t, we define sets M_t and F_t as all males and females, respectively, who married in any county that year. We then randomly sample from the Cartesian product $M_t \times F_t$ to create a set of 1,000 national random marriages per year.

This approach reveals the full extent of potential integration in the absence of both geographic and social constraints. By comparing within-county random marriages to national random marriages, we can isolate the effect of geographic sorting on social integration.

It is important to emphasize that residential location itself represents a choice that affects marriage markets. Individuals can choose where to live and how much to geographically segregate themselves, just as they choose whom to marry within their location. Geographic sorting, therefore, is not an exogenous constraint but rather an endogenous component of social integration strategies. When people choose to live in communities with others similar to themselves, they are already exercising preferences that shape their marriage possibilities before any individual-level spouse selection occurs. This dual nature of sorting—first by location and then by spouse within location—means that the gap between within-county and national random marital distances captures meaningful social preferences rather than merely reflecting fixed geographic constraints.

3.3.3 Measuring Counterfactual Marital Distances

For both sets of random marriages, we calculate marital distances using the same methodology applied to actual marriages. For each counterfactual marriage, we identify the shortest path connecting the husband and wife through the family network that existed prior to the marriage year. This ensures that our comparison treats actual and counterfactual marriages consistently.

An important consideration is that not all random pairs can be connected through pre-existing family ties. For these unconnected pairs, we handle them using the same approaches described in Section 3.4. The primary analysis excludes unconnected pairs, while robustness checks either assign them a maximum distance or use surname-based imputation.

Throughout the period 1750-1950, actual marital distances are consistently shorter than random distances, indicating that people chose spouses who were closer in the social network than would occur by chance. However, the gap between actual and random distances narrowed substantially over time, suggesting a weakening of preferences for socially proximate spouses and greater integration across social boundaries.

3.3.4 Computational Implementation

Our implementation constructs historical subgraphs for each analysis year, containing only family relationships formed before that year. For each year from 1750 to 1950, we:

1. Create a historical subgraph containing all family relationships (both parent-child and marriage edges) formed prior to the analysis year.

2. Identify actual marriages that occurred in that year, recording the husband, wife, and county.

3. Generate within-county random marriages by randomly pairing males and females who married in the same county-year, excluding actual pairings.

4. Generate national random marriages by randomly pairing males and females who married anywhere in that year.

5. For each set of marriages (actual, within-county random, and national random), analyze connectivity in the historical subgraph and measure marital distances.

This approach ensures that our random benchmarks accurately reflect what marriage patterns would look like in the absence of social network preferences, while still respecting fundamental demographic constraints of the historical marriage market.

3.4 Random Marital Distance as a Benchmark

To interpret observed marital distances, we need a benchmark for comparison. We construct this benchmark by randomly pairing individuals who married in the same year. These counterfactual marriages represent what marital distances would be if individuals selected spouses without regard to their position in the social network. The comparison between actual and random marital distances reveals the extent to which marriage choices reflect social structure beyond what would occur by chance.

We implement two complementary versions of random matching. First, we construct within-county random matches, pairing individuals who married in the same county-year. This approach accounts for geographic constraints and local marriage markets, capturing the extent of social sorting conditional on location. Second, we create counterfactual marriages from the national pool of individuals who married in the same year. This approach reveals the full extent of potential integration in the absence of both geographic and social constraints.

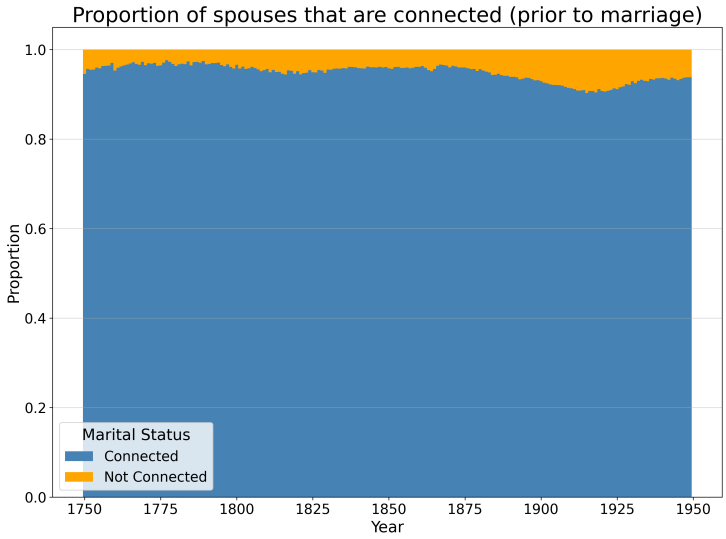
[Emphasize that people can choose where to live - can choose how much to segregate or not geographically.]

Throughout the period 1750-1950, actual marital distances are consistently shorter than random distances, indicating that people chose spouses who were closer in the social network than would occur by chance. However, the gap between actual and random distances narrowed substantially over time, suggesting a weakening of preferences for socially proximate spouses and greater integration across social boundaries.

3.5 Unconnected Spouses

Not all marriages in our sample can be traced through pre-existing family connections. In some cases, no path exists between the spouses in the pre-marriage network, meaning they had no identifiable family link before marriage. These unconnected pairs represent marriages

Figure 4: Proportion of spouses connected before marriage (1750–1950)



between socially distant families whose networks had not previously overlapped. Figure 4 shows the proportion of spouses connected before marriage from 1750 to 1950. For most of this period, nearly all couples shared at least one indirect tie before marriage.

We address this issue using three complementary approaches. Our primary approach is to exclude unconnected couples, as their network distance is undefined within our framework. This ensures that all reported distances reflect actual paths in the family network.

Second, for robustness, we assign unconnected couples a distance of 30, which corresponds to the 99th percentile of observed distances in our connected sample. This approach allows us to include all marriages in our analysis while acknowledging that unconnected couples likely represent the upper tail of the social distance distribution.

Our third approach (in progress) is to develop a surname-based distance imputation method that leverages the structure of marriage patterns between family names.

This proceeds in three steps. First, we construct a weighted surname marriage network, where nodes represent surnames and edges represent marriages between individuals with those surnames. The edge weight between any two surnames (s_i, s_j) is calculated as:

$$w_{ij} = \frac{n_{ij}}{\sqrt{n_i \cdot n_j}} \tag{1}$$

where n_{ij} is the number of marriages between individuals with surnames s_i and s_j, and n_i and n_j are the total number of individuals with these surnames in our marriage data. This normalization prevents common surnames from having artificially strong connections merely due to their frequency.

Second, we calculate the shortest path length through this weighted network for each surname pair, which we call $\text{SurnameDistance}_{ij}$. When finding shortest paths, we transform the weights so that frequently intermarrying surnames (with higher w_{ij} values) are considered "closer" in the network space. This creates a measure of social distance based on marriage patterns between family names.

Third, we build a predictive model using the subset of couples who are connected in the family network. For these couples, we can observe both their actual marital distance and their surname distance, allowing us to estimate:

$$\text{MaritalDistance}_{ijk} = \alpha + \beta \cdot \text{SurnameDistance}_{ij} + \gamma X_k + \varepsilon_{ijk} \tag{2}$$

where $\text{MaritalDistance}_{ijk}$ is the actual observed network distance for couple k with husband's surname i and wife's surname j, $\text{SurnameDistance}_{ij}$ is the surname network distance, and X_k includes additional predictors such as geographic proximity, time period, and regional factors. Using this fitted model, we then impute the most probable network distances for unconnected couples based on their surnames, geographic locations, and other characteristics.

To illustrate this method, consider a simplified example with just three surnames: Smith, Miller, and Garcia. Based on marriage records, we observe:

- Smith-Miller marriages occur at 3× the rate expected by random chance
- Miller-Garcia marriages occur at 2× expected rate
- Smith-Garcia marriages occur at 0.5× expected rate (half the expected rate)

When finding shortest paths, we transform these weights by taking their reciprocals, so higher intermarriage rates become shorter distances:

- Smith-Miller distance: $\frac{1}{3} = 0.33$
- Miller-Garcia distance: $\frac{1}{2} = 0.5$
- Smith-Garcia distance: $\frac{1}{0.5} = 2$

In this network, the shortest path from Smith to Garcia would pass through Miller (Smith $\rightarrow$ Miller $\rightarrow$ Garcia) with a total distance of $0.33 + 0.5 = 0.83$, rather than directly (Smith $\rightarrow$ Garcia) with a distance of 2. This reflects that families with Smith and Garcia surnames are more socially connected through Miller families than directly.

For a Smith-Garcia couple where we cannot observe a family connection, we would use this surname distance along with geographic proximity and time period to estimate their likely genealogical distance. If our model shows that couples with these characteristics typically have network distances around 10 steps, we would impute 10 as the estimated marital distance for this Smith-Garcia couple.

Importantly, our core findings regarding trends in marital distance remain consistent across all three approaches—excluding unconnected couples, assigning them a maximum distance, or using surname-based imputation. This consistency suggests that our results are robust to different methodological choices for handling marriages that span the boundaries of our observable family network.

3.6 Validating Marital Distance as a Measure of Social Integration

To establish marital distance as a meaningful measure of social integration, we examine its relationship with other established metrics of social cohesion and community structure. Building on recent work measuring close-knit communities (Raz, 2025), we validate our marital distance metric against multiple indicators of social structure, both historical and contemporary.

3.6.1 Historical Measures of Integration

In Table 2, we present correlations between marital distance and various measures of social integration at the county-by-decade level (from 1850-1940). The social integration measures are derived from decennial censuses. To align the timeframes, we compute the average marital distance at the county-year level over the ten years following each census and assign it to the corresponding decade.[2] We estimate regressions with marital distance as the independent variable and the various measures (standardized versions) as the outcome variables, while controlling for county and decade fixed effects. For contemporary measures, we show correlations with marital distance from 1940.

First, we find that counties with higher marital distances (indicating tighter social networks) have significantly lower Local Name Index (LNI) scores. Following Fryer and Levitt (2004), the LNI measures in-group identity through naming patterns, quantifying the degree to which parents choose distinctively local names for their children. Formally defined as:

$$\text{LNI}_{nlgt} = 100 \times \frac{\Pr(n|l,g,t)}{\Pr(n|l,g,t) + \Pr(n|-l,g,t)} \tag{3}$$

where n represents a particular first name, l is the geographical level defined as "local," $-l$ refers to all non-local locations, g is gender, and t is the census year. Higher LNI scores indicate stronger communal identification, which aligns with the shorter marital distances observed in more tightly-knit communities. The 10th percentile of marital distance over this period in our dataset is 11, whereas the 90th percentile is 19. Therefore, these results suggest that going from the 10th to the 90th percentile of martial distance (which is 8 steps) reduces a county's LNI score by 0.64σ ($p < 0.001$).

Second, marital distance correlates negatively with Intra-Community Marriage (ICM) rates—the county-level share of married couples with a common birthplace. This measure follows from the research of Triandis (1995), who established that when individuals are bound in close-knit social networks, group membership plays a central role in shaping their social relationships. Like the LNI index, the relationship is statistically significant – going from the 10th to the 90th percentile of martial distance is associated with a reduction of 1.2σ in the county's ICM rate ($p < 0.001$).

Third, marital distance also shows strong negative correlations with measures of cultural tightness, including the Tight Norms Index (TNI) and Religious Homogeneity Index (RHI). Following Gelfand et al. (2006), the TNI measures conformity in family-related practices (mother's age at first birth, number of children, and household composition), while the RHI

[2]For example, the 1850 decade for a given county reflects the average marital distance observed in that county between 1850 and 1859.

captures concentration of religious denominations within a community using data from Manson et al. (2020). These indicators reveal how marital distance relates to broader patterns of social conformity and cultural homogeneity.

Finally, we observe robust correlations between marital distance and measures of kinship tightness introduced by Nelson (2020), including Median Kinship Propinquity (MKP) and Kinship Propinquity Rate (KPR), as well as the Strength of Family Ties Index (SFTI) derived from Ruggles et al. (2020) census data. These relationships hold both across and within states (i.e., both with and without state fixed effects, results not reported), suggesting that marital distance captures meaningful variation in social structure beyond regional differences.

Table 2: Correlates of Marital Distance

Outcome	Coefficient	Std. Error	p-value	N
Panel A: Historical Measures				
Local Name Index (LNI)	-0.082***	0.003	< 0.001	21,624
Intra-Community Marriage Rate	-0.151***	0.004	< 0.001	21,632
Kinship Propinquity Rate (KPR)	-0.156***	0.004	< 0.001	21,666
Median Kin Propinquity (MKP)	-0.100***	0.004	< 0.001	21,635
Tight Norms Index (TNI)	-0.055***	0.003	< 0.001	21,602
Strength of Family Ties Index (SFTI)	-0.101***	0.003	< 0.001	20,074
Religious Homogeneity Index (RHI)	-0.042***	0.004	< 0.001	18,949
Panel B: Contemporary Measures				
Social Clustering (Facebook, 2022)	-0.031***	0.005	< 0.001	2,537
Social Support Ratio (Facebook, 2022)	-0.017***	0.006	0.001	2,537
Communal vs. universal values (2008-2018)	-0.002	0.006	0.744	1,906

* p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the county level in Panel A, while robust standard errors are reported in Panel B. In Panel A, the reported coefficients come from regressions of various outcome variables on marital distance at the county-by-decade level. In Panel B, contemporary county-level measures are regressed on average marital distance from the 1940s. All of the outcome variables are standardized, but marital distance is not.

3.6.2 Contemporary Measures of Integration

Marital distances from the 1940s correlate significantly with contemporary measures of social capital collected decades later. We examine two county-level indicators of local close-knit social networks derived from Facebook data by Chetty et al. (2022): (i) Social Clustering, which measures the average fraction of an individual's friend pairs who are also friends with each other, and (ii) Social Support Ratio, which measures the proportion of within-county friendships where the pair of friends share a third mutual friend within the same county. These data are calculated from individuals aged 25-44 residing in the US in 2022.

Panel B of Table 2 shows that counties with shorter historical marital distances exhibit higher Social Clustering and Social Support Ratio scores today – indicating greater network density and cohesion.

Beyond network structure, one might expect historical marital distances to correlate with contemporary cultural and psychological traits associated with close-knit communities. Following Enke (2020), we examine county-level measures of moral values derived from the Moral Foundations Questionnaire developed by Graham et al. (2011), focusing on the relative importance of communal versus universalist moral foundations. Surprisingly however, we do not find a strong association between these two measures.

In sum, through these multiple validation exercises, we establish marital distance as a robust and meaningful measure of social integration that captures fundamental aspects of community structure across time and space. The strong correlations with both historical and contemporary indicators suggest that this network-based approach provides valuable insights into the long-term evolution of American social integration.

The primary advantage of marital distance over conventional group-based measures used here is its comprehensive scope and coverage. Our measure spans a long period (1750-1950), covers a wide geographic area, offers annual frequency, provides individual-level variation, and allows for location-ally disaggregated analysis. This granularity enables us to trace the evolution of social integration with unprecedented temporal and spatial precision.

4 Patterns in Marital Distance

This section documents key empirical patterns in marital distance across time and space. We show that marital distance increased substantially between 1750 and 1950, with significant geographic variation in both levels and trends. These patterns provide evidence of a fundamental transformation in American social integration coinciding with the country's economic development.

Figure 5 compares the distribution of actual marital distances to those from random pairings. The actual distribution is centered around lower values (peaking at 12-13 steps) compared to the random distribution (peaking at 17 steps). This shift indicates a systematic preference for socially proximate spouses.

4.1 Changes Over Time

The network distance between spouses changed substantially between 1750 and 1950. Figure 6 shows that average marital distance rose from approximately 9.5 steps in 1750 to 16 steps by 1950, with the sharpest increase occurring between 1870 and 1915. This trend might reflect two distinct forces: changes in the underlying social network structure as the U.S. population grew and evolved, or changes in how people chose their spouses within that network.

To distinguish between these possibilities, we construct a random matching benchmark. For each year, we calculate what marital distances would have been if men and women who married that year had been randomly matched. Figure 7 compares actual marital distances to this random benchmark, revealing several key patterns. Throughout the period, actual marital distances are far shorter than random distances, indicating that people consistently chose spouses who were closer in the social network than would occur by chance. However,

Figure 5: Distribution of marital distances

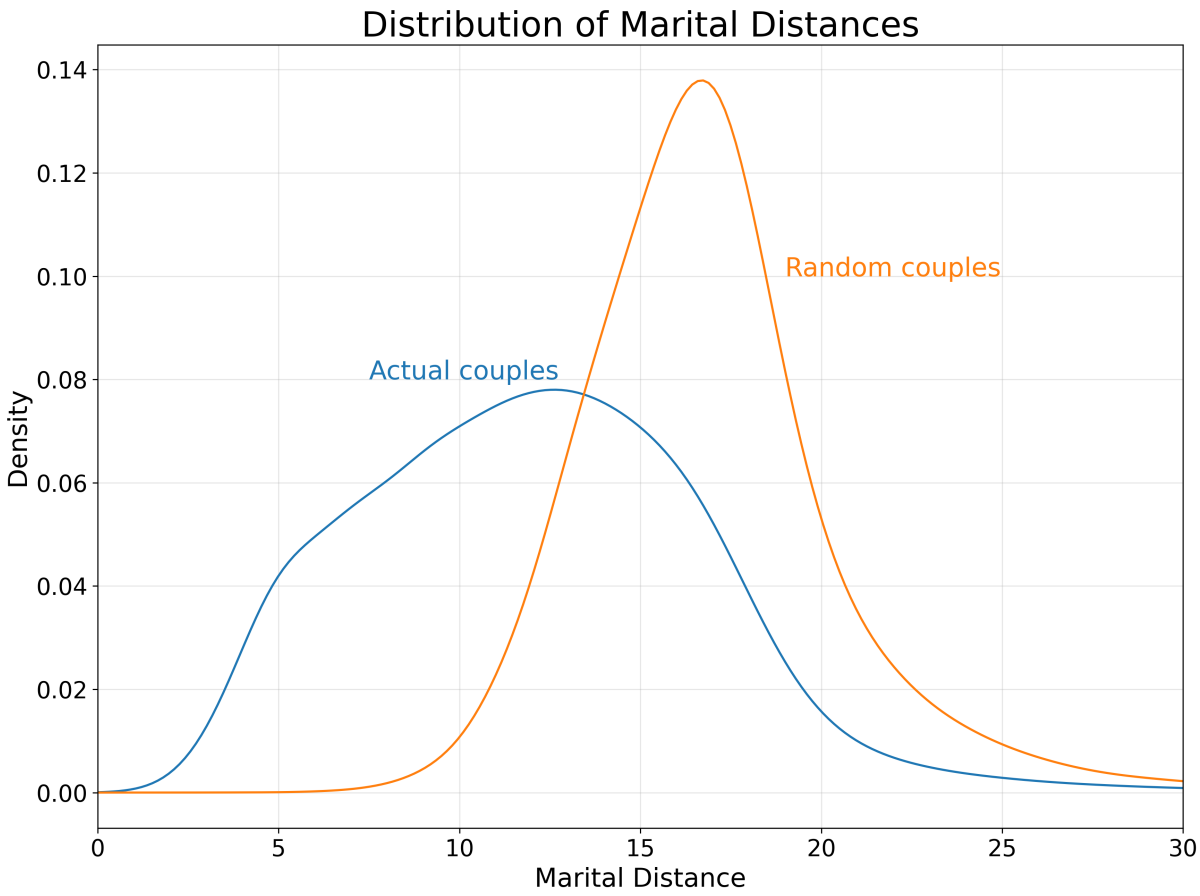


Figure 6: Marital distance over time (1750–1950)

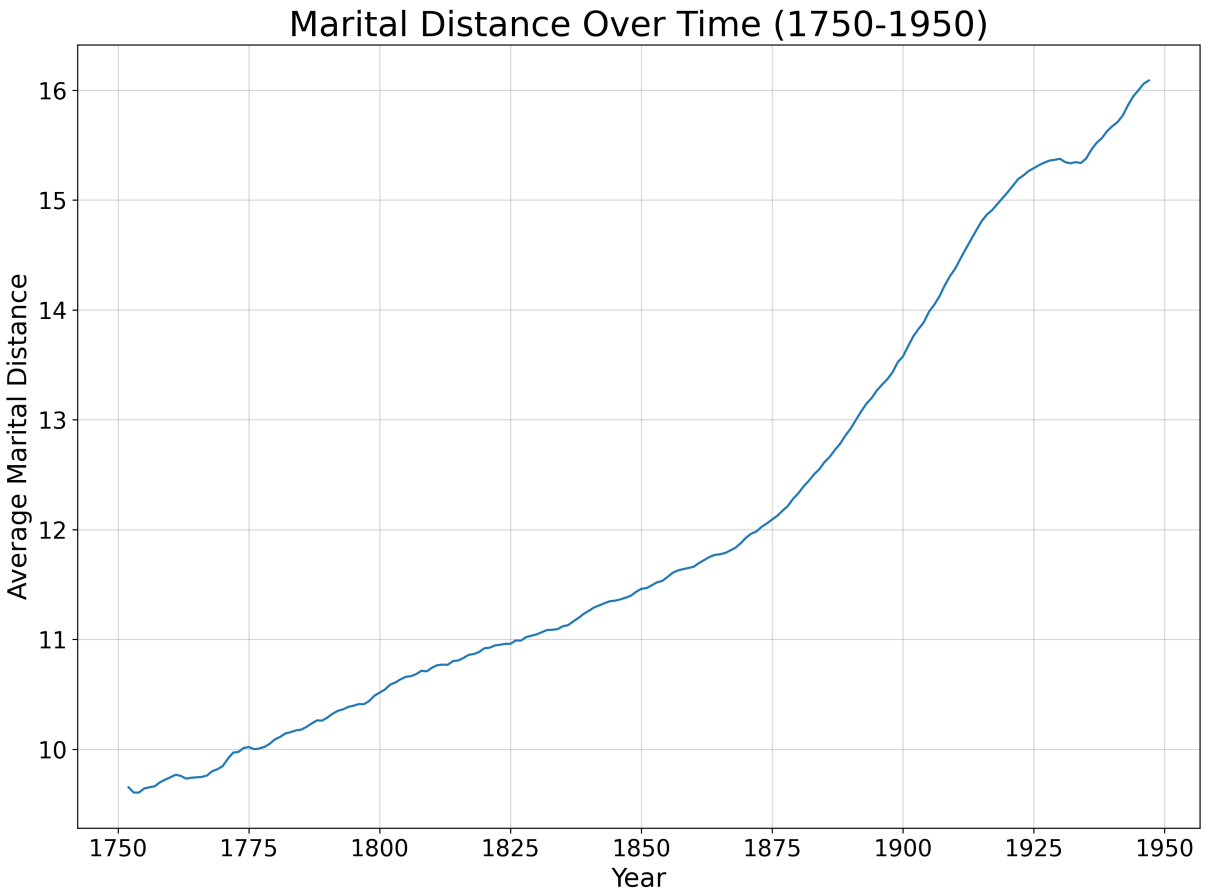
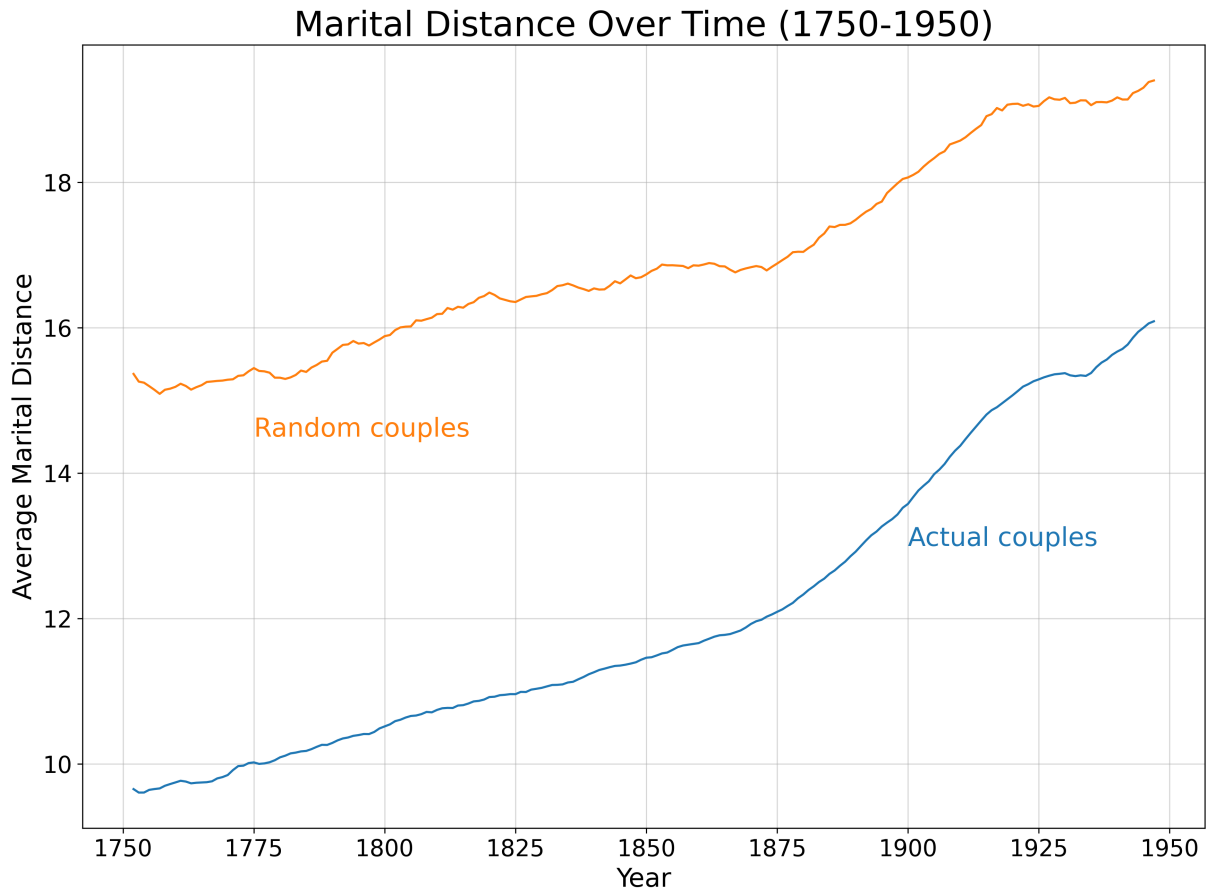


Figure 7: Marital distance over time (1750–1950)



the gap between actual and random distances narrowed substantially over time, with the most pronounced convergence occurring after 1870.

This convergence is particularly meaningful because it occurred despite the rising random benchmark. As shown in Table 3, random marital distances increased at a rate of 0.17 steps per decade, reflecting how population growth and mobility expanded the network and pushed potential matches further apart. Yet actual marital distances increased much faster, at 0.37 steps per decade. A community determined to maintain close marriage ties could have continued to do so even as the network grew—for instance, by consistently choosing spouses within a fixed social distance regardless of broader demographic changes. Instead, marriage patterns evolved in a way that suggests people became increasingly willing to marry outside their immediate social circles.

The pattern of increasing social distance through marriage becomes even clearer when examining the prevalence of close-kin marriages. Figure 8 shows that the share of marriages occurring within short marital distances declined dramatically from 1750 to 1950. The proportion of marriages within a marital distance of 10 fell from over 65% in 1750 to about 10% by 1950, with similar declines for marriages within distances of 8, 6, and 4. This suggests a long-term shift away from kin-based marriage patterns toward unions that bridged more

Table 3: Time Trends in Marital Distance by Category (1750-1950)

	(1) Actual Couples	(2) Random (county-year)	(3) Random (year)
year	0.037***	0.031***	0.017***
	(0.000)	(0.000)	(0.000)
Observations	3046885	4293122	134581
Adjusted R2	0.105	0.081	0.037

Standard errors in parentheses. Coefficients represent annual change in marital distance.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

distant social circles.

4.2 Geographic Variation

Marital distance varied substantially across regions throughout the period we study. Figure 9 shows average marital distance by region from 1750 to 1950, revealing persistent geographic patterns. The South consistently had shorter marital distances, while the Northeast and Midwest had longer distances. Western states generally had high distances once settled, with a rapid rise after 1850 coinciding with frontier expansion and migration. Despite these differences, all regions experienced increases in marital distance over time, with regional gaps narrowing in later periods.

To examine these patterns at a finer level, we calculate average marital distance by county. Figure 1 reveals substantial variation even within states. For marriages occurring between 1900 and 1949, marital distance was highest in the western and northern United States, while southeastern and Appalachian regions exhibited lower marital distances. This geographic distribution aligns with broader patterns of economic development during the early 20th century.

Table 4 confirms these regional differences in a regression framework. Even after controlling for individual characteristics, the South had marital distances 1.5 steps shorter than the Northeast (column 1). Meanwhile, the West had distances 1.3 steps longer than the Northeast, with the Midwest falling between these extremes at 0.3 steps above the Northeast baseline. These differences persist with year fixed effects (column 2).

4.3 Individual and Family Correlates of Marital Distance

Marital distance varies systematically with individual and family characteristics. Table 4 presents regression results examining these relationships while controlling for time period and location fixed effects. Several key patterns emerge.

Geographic mobility strongly predicts greater marital distance. Individuals who moved between states before marriage have marital distances 1.6 steps longer than non-movers, while those whose parents moved states before their birth have distances 0.6 steps longer. These effects are additive and remain robust across all specifications, indicating that geographic mobility is strongly associated with marrying outside one's close social network.

Figure 8: Decline in close-kin marriage over time (1750–1949)

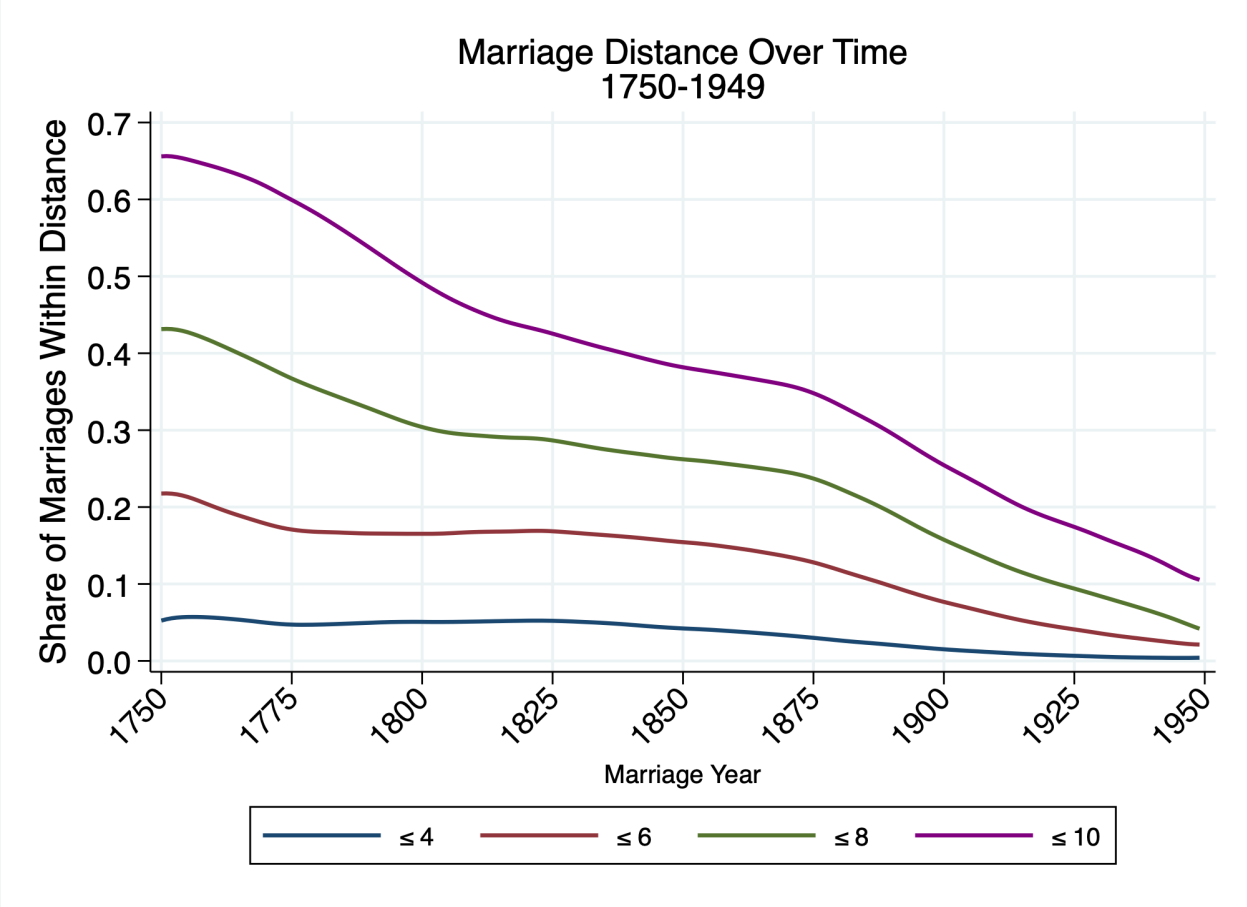


Figure 9: Marital distance over time by region (1750–1950).

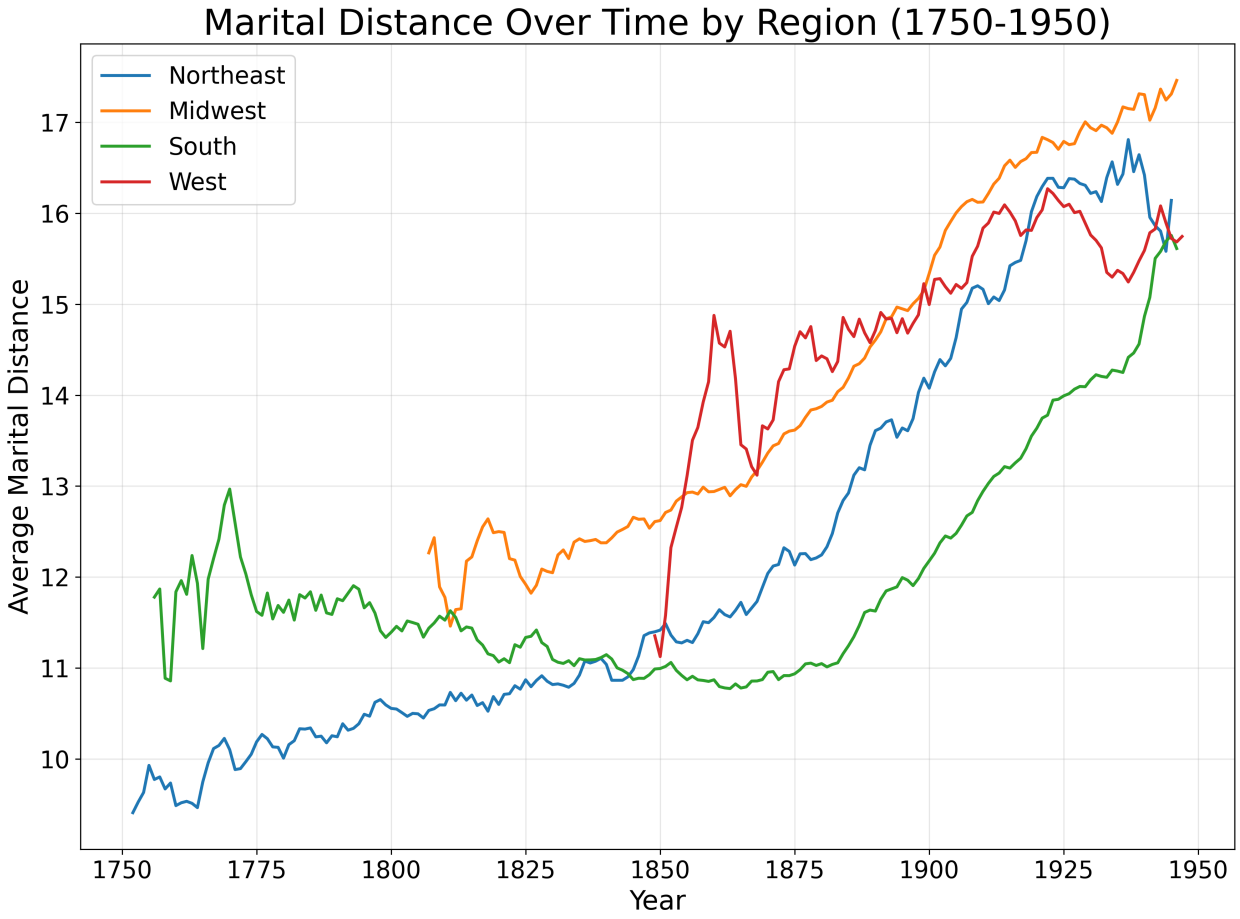


Table 4: Determinants of Marital Distance

	(1)	(2)	(3)	(4)
Parents moved states pre-birth	0.607***	0.649***	0.535***	0.525***
	(0.026)	(0.025)	(0.025)	(0.027)
Moved states before marriage	1.557***	1.581***	1.286***	1.247***
	(0.020)	(0.020)	(0.025)	(0.026)
2nd-gen Immig (1 parent)	2.926***	2.784***	2.750***	2.701***
	(0.077)	(0.052)	(0.052)	(0.054)
2nd-gen Immig (both parents)	7.564***	7.379***	7.196***	7.242***
	(0.165)	(0.087)	(0.086)	(0.090)
Wife's age at marriage	-0.017***	-0.017***	-0.014***	-0.015***
	(0.002)	(0.002)	(0.002)	(0.002)
Husband's age at marriage	-0.001	0.004**	0.001	0.001
	(0.002)	(0.001)	(0.001)	(0.002)
Midwest	0.345***	0.505***		
	(0.028)	(0.026)		
South	-1.498***	-1.318***		
	(0.024)	(0.022)		
West	1.332***	1.293***		
	(0.074)	(0.065)		
Marriage year	0.031***			
	(0.000)			
Observations	261782	261782	261781	238475
Adjusted R2	0.167	0.178	0.204	0.217
Year FE	No	Yes	Yes	Yes
State FE	No	No	Yes	Yes
County FE	No	No	No	Yes

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Age at marriage shows a small but statistically significant relationship with marital distance. Wife's age at marriage is negatively associated with marital distance, with each additional year corresponding to a 0.02 step decrease. This pattern may reflect different marriage market dynamics for women marrying at older ages, though the magnitude of the effect is relatively small compared to geographic mobility and immigration status.

Immigration status shows a strong relationship with marital distance. Second-generation immigrants with both parents born abroad exhibit marital distances 7.6 steps longer than native-born Americans with native-born parents. Those with one foreign-born parent have distances 2.9 steps longer. These substantial differences suggest that children of immigrants occupied a distinct position in American social networks, maintaining some distance from established native networks while not fully assimilating through marriage.

One potential concern with our marital distance measure is that it may be artificially inflated for immigrants due to missing genealogical data from their countries of origin. Given that the period of sharpest increase in marital distance (1870-1915) coincides with high immigration rates, this could confound our interpretation of increasing social integration. To address this concern, we repeat our main analyses on a sample restricted to individuals with parents and grandparents all born in the United States. As shown in Figure X, the temporal pattern of increasing marital distance remains robust in this native-only sample, though the magnitude of increase is somewhat attenuated. Furthermore, when we control for county-level immigration rates in our manufacturing analysis (Table Y), the relationship between manufacturing employment and marital distance remains significant, suggesting that our findings reflect genuine changes in social integration beyond compositional effects of immigration.

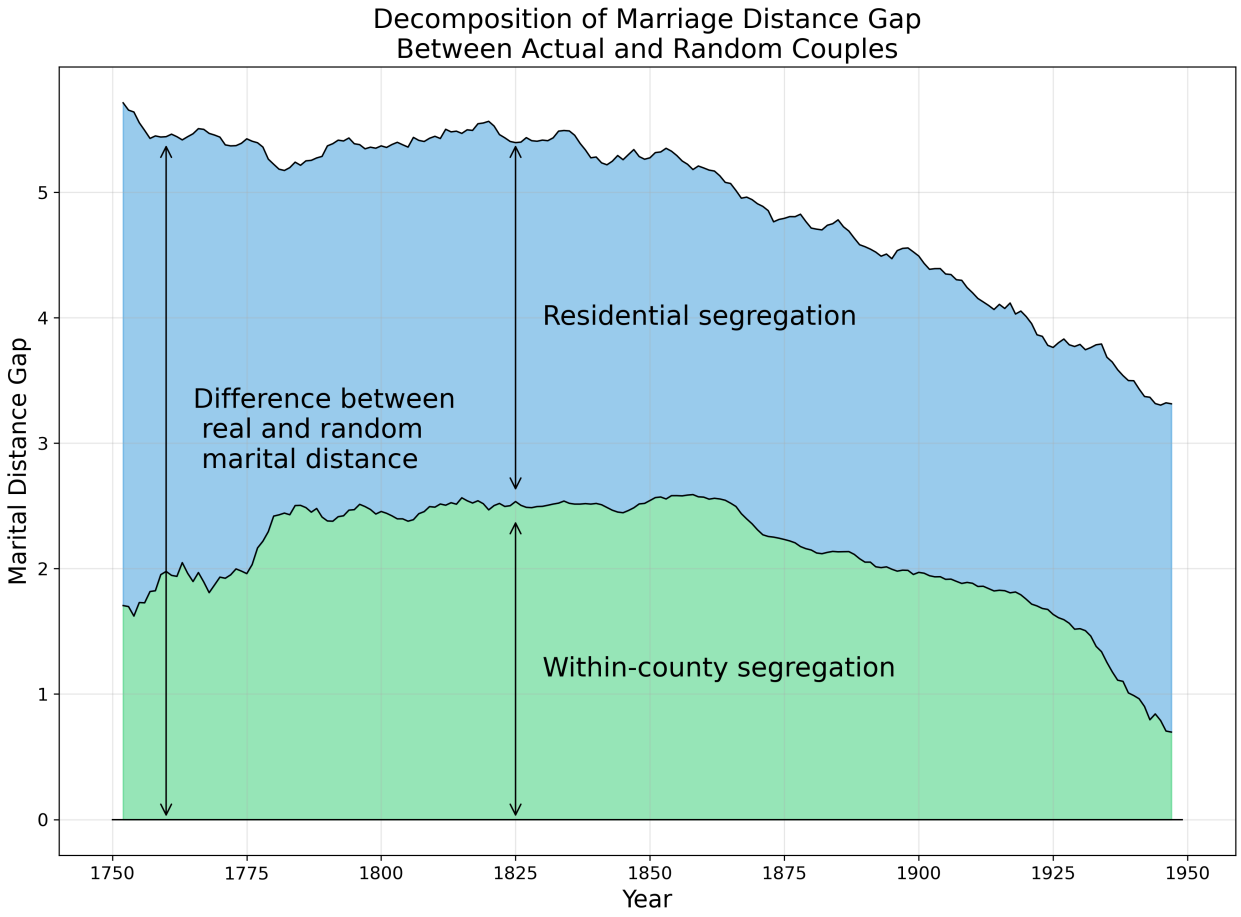
4.4 Decomposing the Gap Between Actual and Random Marital Distance

To better understand the constraints on integration, we decompose the gap between actual and random marital distances into within-county and between-county components. Figure 10 shows that the total gap remained relatively stable from 1750 to 1900 before declining in the early 20th century. Within-county segregation accounts for a substantial portion of this gap, reflecting strong kinship and local marriage preferences. The persistence of this gap suggests that while marriages became more exogamous over time, they remained more endogamous than expected under random pairing.

This decomposition provides important context for interpreting the relationship between industrialization and social integration. The narrowing gap after 1900 suggests increasing openness in marriage patterns, likely due to urbanization and mobility associated with industrial development. However, the persistence of a substantial gap even in the 20th century indicates that social constraints on marriage remained meaningful despite economic transformation.

This pattern implies that while industrialization weakened local kinship-based marriage constraints, broader geographic sorting continued to influence social integration. As kinship ties became less determinative of marriage choices within communities, the persistent difference between within-county and national random marriage patterns suggests that the local

Figure 10: Decomposition of marriage distance gap between actual and random couples



Note: The gap between actual and random marital distances is decomposed into two components: within-county segregation (green) and across-county residential segregation (blue).

nature of marriage markets remained an important factor in structuring social integration well into the 20th century.

5 Manufacturing and Social Integration

A central question in understanding social transformation is whether economic development actively reshaped patterns of social integration. Industrialization fundamentally altered the economic landscape by restructuring where people lived, how they worked, and with whom they interacted, but did it substantively change the underlying social fabric connecting different groups? Did manufacturing growth facilitate greater social connectivity across previously separate communities? If so, we should find that industrialization increased network distances between spouses, indicating that economic modernization actively dissolved traditional social boundaries.

In this section, we test this hypothesis by examining whether the spread of manufacturing increased marital distance. We begin with county-level analysis, exploiting variation in the timing and intensity of industrialization across locations. The results suggest that increased manufacturing employment led to substantial increases in marital distance, providing evidence that industrialization reshaped social networks.

5.1 County-Level Analysis

5.1.1 Panel Specification

Our baseline specification examines how changes in manufacturing employment within counties affect marital distance:

$$\text{MaritalDistance}_{ict} = \beta \times \text{ManufEmp}_{ct} + \gamma X_{ct} + \alpha_c + \delta_t + \varepsilon_{ict}$$

where i indexes marriages, c indexes counties, and t indexes decades. The dependent variable $\text{MaritalDistance}_{ict}$ is the network distance between spouses at marriage. The key independent variable ManufEmp_{ct} is manufacturing employment as a share of county population. We include county fixed effects α_c to control for time-invariant county characteristics and year fixed effects δ_t to account for national trends. The control vector X_{ct} includes county population, average random marital distance, urban population share, and infrastructure measures like railroad and waterway access.

Standard errors are clustered at the county level to account for serial correlation in the error terms. We restrict the sample to marriages between 1850-1940 (years with manufacturing data), and couples where both spouses are native-born and have all eight grandparents (four for each spouse) identified in our data.

5.1.2 Instrumental Variables Strategy

A key concern is that manufacturing location decisions may respond to local social conditions. To address this endogeneity, we construct a shift-share instrument that isolates plausibly exogenous variation in local manufacturing growth (Borusyak et al., 2022). For each county c, we calculate:

$$Z_{ct} = \sum_j \theta_{cj0} \times g_{jt}$$

where θ_{cj0} is industry j's initial employment share in county c, and g_{jt} is the leave-one-out national growth rate in industry j's employment share. This provides variation in manufacturing exposure that depends only on initial industrial composition and national trends.

The instrument's validity rests on two key assumptions: (1) initial industry shares are not correlated with trends in marriage patterns and (2) national industry growth rates are not driven by county-specific marriage trends. We evaluate these assumptions through pre-trend tests, robustness to alternative industry share measures, and tests for correlation between initial shares and marriage patterns.

5.2 Main Results

Table 5 presents our main results. Column 1 shows the baseline panel specification, where a ten percentage point increase in manufacturing employment share is associated with a 0.8 step increase in marital distance. This relationship remains stable when controlling for county population in Column 2. Population serves as a proxy for urbanization, though we recognize this is an imperfect measure; future versions will incorporate more direct urbanization metrics to better isolate the manufacturing effect from broader urban development patterns.

To address endogeneity concerns, Columns 4 and 5 introduce an instrumental variables (IV) approach using shift-share variation in national industry trends. The IV estimates are substantially larger, suggesting that OLS may understate the causal effect of manufacturing on social integration. A ten percentage point increase in manufacturing employment raises marital distance by approximately 1.2 steps, indicating that industrialization played a significant role in reshaping marriage patterns.

Columns 3 and 6 help disentangle the mechanisms behind this effect by controlling for the county's average random marital distance. This adjustment separates two distinct pathways through which manufacturing increased marital distances. The results suggest that about three-quarters of the observed increase stems from structural changes in county social networks—not merely the presence of new residents, but the fundamental fragmentation of family networks as manufacturing attracted migrants who became separated from their extended kinship systems. When people moved for industrial work, they typically arrived with only nuclear families or as individuals, leaving behind the dense web of relatives that had traditionally provided marriage partners. This network disruption forced even those with strong endogamous preferences to marry at greater social distances simply because close kin were no longer locally available. The remaining quarter of the effect reflects actual changes in marriage preferences—people becoming more willing to marry outside their close social circles even when traditional partners were available.

This distinction has important implications for understanding social integration during industrialization. Rather than merely reflecting demographic shifts, the large "compositional" effect captures how migration fundamentally restructured American community networks from densely interconnected webs to more fragmented archipelagos of smaller family

Table 5: Effect of Manufacturing employment on marital distance

	(1)	(2)	(3)	(4)	(5)	(6)
Manuf emp share (county)	8.240***	8.343***	2.245***	12.47***	12.73***	3.276***
	(2.147)	(1.920)	(0.353)	(2.563)	(2.419)	(0.656)
Log county population		-0.585	0.0429		-0.599	0.0269
		(0.631)	(0.0700)		(0.476)	(0.0584)
Avg. random marital distance			0.704***			0.689***
			(0.0316)			(0.0264)
Observations	5309	5309	5309	5007	5007	5007
Number of counties	1,741	1,741	1,741	1,439	1,439	1,439
Adjusted R2	0.883	0.968	0.968	0.406	0.841	0.841
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
IV (Shift-share)	No	No	No	Yes	Yes	Yes
First stage F-stat				28.9	28.8	23.4

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Standard errors in parentheses, clustered at the county level. The dependent variable is the average marital distance in a county. "Manuf emp share" is the share of employed individuals reporting manufacturing occupations in each census decade. Marriage observations are assigned to their corresponding census decade's manufacturing data. All specifications include county and year fixed effects. Columns (1)-(3) present OLS estimates, while columns (4)-(6) use a shift-share IV approach with first stage F-statistics reported. Columns (3) and (6) control for county-level random marital distance to account for population composition.

units. Meanwhile, the preference change component suggests industrialization also weakened traditional kinship-based marriage norms, though to a lesser extent. Together, these mechanisms transformed not only where Americans lived and worked, but the very structure of social relationships that had previously constrained marriage possibilities. Manufacturing thus advanced integration both by disrupting traditional network structures and by gradually shifting cultural preferences toward more open marriage patterns.

5.3 Robustness

5.4 Magnitudes and Interpretation

To interpret the magnitudes, we perform several exercises. The estimates suggest manufacturing growth can explain about 40% of the increase in marital distance between 1870 and 1915. Effects are similar across regions but somewhat larger in the Northeast. Impacts are larger in counties with lower initial marital distances, suggesting that industrialization had a particularly strong effect in previously more insular communities. Results are stronger for manufacturing industries employing women, pointing to potential mechanisms through female labor market participation.

5.5 Mechanisms

6 Individual-Level Analysis

In this section, we examine whether marital distance has direct effects on individual economic outcomes, including occupational choice, geographic mobility, and household structure. This analysis helps determine whether the observed relationship between industrialization and social integration merely reflects structural economic changes, or whether social integration through marriage networks actively facilitated economic mobility.

6.1 Empirical Strategy

Estimating the causal effect of marital distance on economic outcomes poses significant empirical challenges. Individuals who marry at greater network distances may differ in unobserved ways that also predict economic success—for example, more ambitious or mobile individuals may both marry outside their immediate social circles and achieve better occupational outcomes. To mitigate these concerns to the extent possible, we include an extensive set of control variables that capture factors influencing both marital preferences and economic outcomes. In addition, we employ a specification with sibling fixed effects, comparing brothers within the same family, which helps to account for family-level selection.

6.1.1 Baseline Specification

Our primary approach is a regression framework that controls for rich demographic and geographic characteristics:

$$y_{ic} = \beta \text{MarDist}_i + X_i\gamma + \alpha_c + \epsilon_{ic}$$

where y_{ic} is an economic outcome for individual i from birth cohort c, MarDist_i is their marital distance, X_i is a vector of controls, and α_c represents birth cohort fixed effects. The control set includes age at marriage, birth state × birth decade interactions, parental socioeconomic background, and county-level manufacturing employment at the time of marriage. Standard errors are clustered at the birth county level to account for spatial correlation in marriage and labor market patterns.

6.1.2 Sibling Fixed Effects

To further mitigate selection concerns, we implement a sibling fixed effects strategy, comparing brothers who marry at different network distances:

$$y_{ifc} = \beta \text{MarDist}_{if} + X_{if}\gamma + \alpha_f + \epsilon_{ifc}$$

where f indexes families, and α_f represents family fixed effects. This specification controls for all shared family characteristics, such as parental resources, cultural norms, and geographic constraints, isolating variation in marital distance that occurs within families.

While sibling fixed effects address many sources of bias, they do not account for selection within families. If more ambitious siblings systematically marry at greater distances, the estimates may still be upward biased. We explore this concern by examining pre-marriage characteristics to assess differences between siblings, testing whether marital distance 'predicts' pre-marital characteristics within sets of siblings.

6.2 Effects on Economic Outcomes

Figure 11 shows the relationship between marital distance and key economic outcomes, highlighting systematic shifts in residence, occupational status, geographic mobility, and household composition.

Individuals with low marital distance (around 5) commonly reside on farms ($\sim$ 75%), while those with high marital distance (around 20) show a lower rate ($\sim$ 60%), a 15 percentage point decline. The opposite pattern holds for urban residence, which increases from about 6% to 14% as marital distance rises. Occupational status also increases with marital distance: occupational income scores rise from approximately 16 to 20, representing a 25% increase. These scores measure median earnings by occupation in hundreds of 1950 dollars, with typical values ranging from 9 for farmers to 40+ for professionals. Consistent with these patterns, individuals with higher marital distance tend to reside in more populated areas, reflecting their integration into more urbanized settings.

Marriage patterns also predict higher geographic mobility. The probability of moving to a different county after marriage rises from 52% to 62% as marital distance increases, while co-residence with parents or in-laws declines from about 26% to 16%. These results suggest that marrying outside close kin networks is associated with greater economic independence and long-term mobility.

Figure 11: Post-marriage census outcomes: Residence, income, migration, household composition

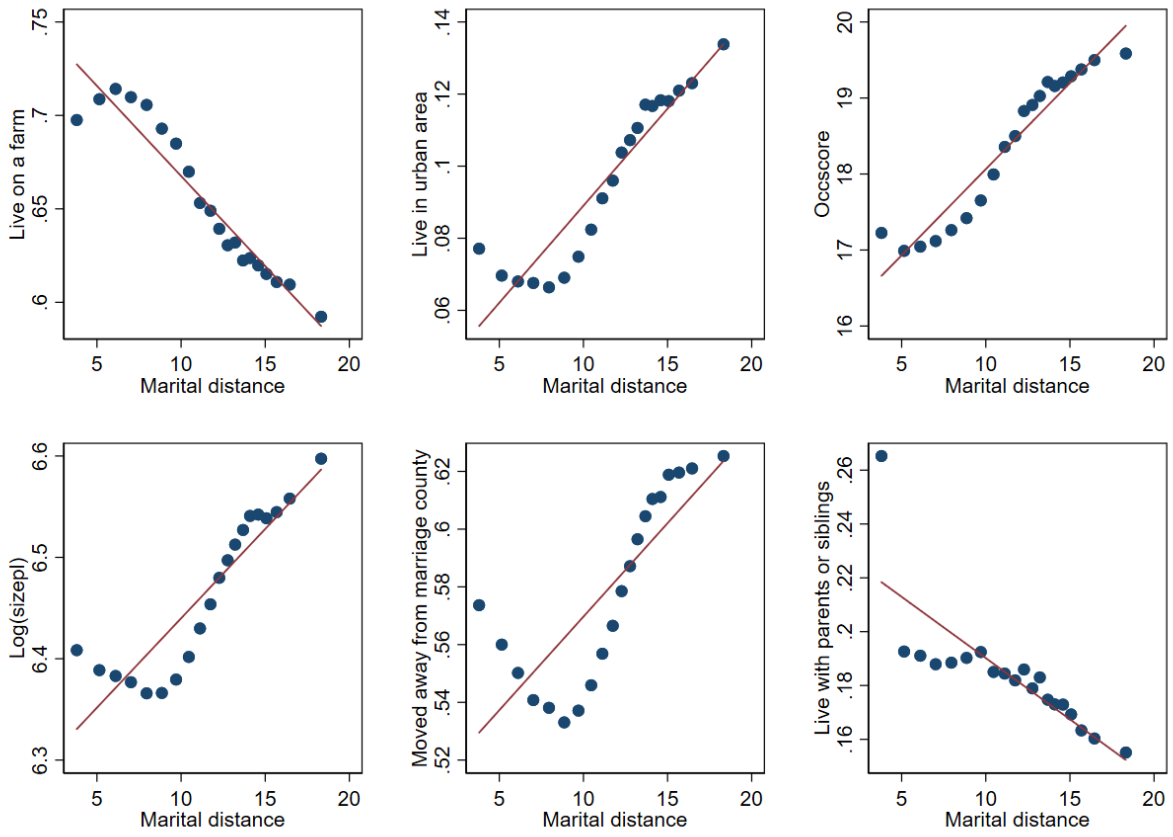


Figure 12: Post-marriage census outcome: Occupational category

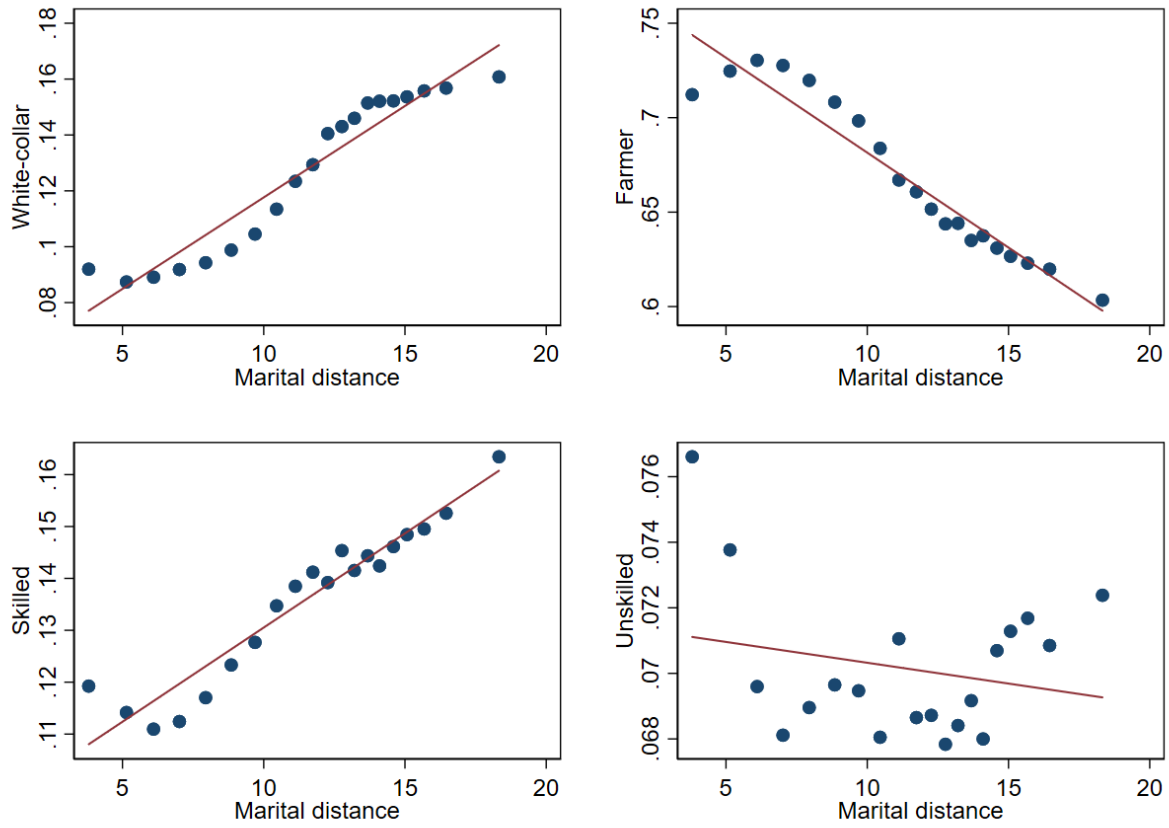


Figure 12 further decomposes occupational transitions, illustrating how employment shifts away from farming and toward white-collar and skilled occupations as marital distance increases. Farming declines from 75% at the lowest marital distances to 60% at the highest, while white-collar employment increases from 8% to 18%. Skilled employment follows a similar pattern, rising from 11% to 16%. In contrast, unskilled employment remains largely unchanged, fluctuating around 7% with no clear trend.

These patterns suggest that individuals who marry farther in the network tend to transition into higher-status occupations, particularly white-collar and skilled jobs, while reducing their likelihood of remaining in farming. This shift aligns with the hypothesis that greater social integration through marriage facilitated economic mobility by broadening access to labor markets and weakening reliance on family-based occupational inheritance.

6.3 Sibling fixed effects

Table 6 presents the estimated effects of marital distance on economic outcomes, complementing the patterns observed in Figures 11 and 12. The baseline specifications show strong associations between greater marital distance and shifts in residence, occupation, and mobility. Including sibling fixed effects attenuates the estimates but preserves the core patterns,

suggesting that these effects are not purely driven by family-level selection.

In the baseline specification, a one standard deviation increase in marital distance is associated with a 9.7 percentage point decrease in farm residence and a 5.4 percentage point increase in urban residence. These effects shrink to 4.0 and 1.7 percentage points, respectively, with sibling fixed effects. Occupational income scores increase by 2.27 points (approximately 12% of the sample mean), with a reduced but still significant effect of 0.886 points under sibling fixed effects.

Marital distance is also linked to increased geographic mobility. The probability of moving to a different county after marriage rises by 6.5 percentage points in the baseline model and 1.7 percentage points with sibling fixed effects. Similarly, individuals with greater marital distance tend to live in larger population centers, with log population size increasing by 0.176 in the baseline model and 0.051 with sibling fixed effects.

Household composition also shifts with marital distance. The likelihood of co-residing with parents or in-laws decreases by 4.6 percentage points in the baseline model and 3.3 percentage points with sibling fixed effects. These patterns suggest that marrying at greater social distances is associated with greater residential and economic independence.

Table 6: Post-marriage census outcomes: sibling fixed effects

	(1) Farm	(2) Urban	(3) Occscore	(4) Log(sizepl)	(5) Moved to Diff. County	(6) Live with parents(-in-law)
Baseline	-.097***	.054***	2.27***	.176***	.065***	-.046***
	(.001)	(.0006)	(.0267)	(.0023)	(.0012)	(.001)
With Sib. FE	-.04***	.017***	.886***	.051***	.017***	-.033***
	(.0018)	(.0009)	(.0491)	(.0034)	(.0026)	(.0019)
Sample mean	.654	.097	18.382	6.465	.579	.184
N	1,874,304	1,874,304	1,874,304	1,874,304	1,874,304	1,874,304

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Standard errors in parentheses. Each column shows coefficients from two separate regressions of the outcome on standardized marital distance. All specifications include fixed effects for age groups (decadal) and birth state × birth decade. "With Sib. FE" adds sibling fixed effects. Outcomes measured in post-marriage census records: Farm (indicator for living on a farm), Urban (indicator for urban residence), Occscore (occupational income score), Log(sizepl) (log population of place of residence), Moved to Diff. County (indicator for residing in different county than marriage), Live with parents(-in-law) (indicator for co-residence with husband or wife's parents). Sample restricted to men with observed post-marriage census records; individuals may appear multiple times if observed across different censuses.

Table 7 further examines occupational shifts associated with marital distance. Consistent with Figure 12, the results indicate a movement away from farming and toward white-collar and skilled occupations.

In the baseline specification, a one standard deviation increase in marital distance is associated with a 6.6 percentage point increase in white-collar employment, a 10.1 percentage point decrease in farming, and a 3.6 percentage point increase in skilled employment. These effects attenuate under sibling fixed effects but remain significant: 2.6 percentage points for

white-collar work, -4.1 percentage points for farming, and 1.2 percentage points for skilled employment. Unskilled employment remains largely unchanged.

Table 7: Post-marriage census outcomes: sibling fixed effects (Occupational category)

	(1) White-collar	(2) Farmer	(3) Skilled	(4) Unskilled
Baseline	.066***	-.101***	.036***	-.001*
	(.0007)	(.001)	(.0007)	(.0005)
With Sib. FE	.026***	-.041***	.012***	.003**
	(.0013)	(.0019)	(.0013)	(.001)
Sample mean	.127	.667	.136	.07
N	1,789,124	1,789,124	1,789,124	1,789,124

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Standard errors in parentheses. Each column shows coefficients from two separate regressions of the outcome on standardized marital distance. All specifications include fixed effects for age groups (decadal) and birth state × birth decade. "With Sib. FE" adds sibling fixed effects. Outcomes are occupational indicators measured in post-marriage census records. The sample is restricted to men with observed post-marriage census records; individuals may appear multiple times if observed across different censuses.

These findings suggest that greater social integration through marriage facilitated economic mobility by broadening access to labor markets and reducing dependence on traditional family-based occupational inheritance. The combination of residence, income, and occupational shifts indicates that individuals who married outside close kin networks were more likely to transition into urban settings, higher-status occupations, and more economically independent household arrangements.

6.4 Social Integration and Inter-generational Mobility

The flexibility of the U.S. economy has long been linked to its socially integrated, mobile population. Unlike rigid kin-based economies, the American model relied on nuclear households that could migrate for work, adapt to new environments, and form diverse social and economic ties. This adaptability is consistent with the broad marriage networks we observe.

Our previous findings indicate that social integration through marriage actively drove economic mobility. Notably, even among siblings, the relationship between marital distance and economic outcomes persisted, suggesting that the benefits of broader marriage ties extend beyond mere family-level selection. Individuals who formed wider marriage networks were more likely to transition into higher-status occupations, relocate to larger towns, and gain economic independence. As industrialization expanded opportunities, breaking free from traditional kinship constraints enabled access to broader labor markets and new economic niches, fundamentally reshaping American society by transforming social networks.

It is natural to assume then that the effects of integration through marriage would persist across generations. A preference or exposure to marriage outside of one's kin network likely cultivates a familial culture that values independence, mobility, and diverse social ties. This cultural capital is likely transmitted from one generation to the next, influencing educational and occupational trajectories and also geographic mobility. Children growing up in households that embraced broader marriage networks may benefit from a richer mix of social capital, exposure to alternative role models, and a supportive environment that encourages exploration beyond insular community ties.

Table 8: The Effect of Marital Distance on Socio-Economic Outcomes of the Children

	(1) Farm	(2) Urban	(3) Occscore	(4) Log(sizepl)	(5) Live with parents(-in-law)
W/O Sib. FE	-.053***	.055***	1.11***	.272***	0.00
	(.0008)	(.0007)	(.0181)	(.0033)	(0.000)
With Sib. FE	-.02***	.016***	.396***	.083***	-.004***
	(.0014)	(.0012)	(.0348)	(.0054)	(.0007)
Sample mean	.479	.269	20.246	7.36	.112
Observations	6,778,840	6,778,840	6,778,840	6,778,840	6,778,840

Note: This table presents the effect of increasing the marital distance by 10 units on various socio-economic outcomes of the children. The first row ("W/O Sib. FE") displays estimates obtained without parental sibling fixed effects, while the second row ("With Sib. FE") presents estimates obtained with parental sibling fixed effects. The sample for this table consists of U.S.-born males born after 1750 who were aged 18 or older at the time of the census to which their genealogical records are linked. Additional controls include birth decade × birth state fixed effects and age dummies (in 10-year intervals) at the time of the census. * for p<.05, ** for p<.01, and *** for p<.001.

The impact of marital distance on intergenerational residential and economic mobility is shown in Tables 8 and 9. Table 8 shows the relationship between marital distance and children's residential and income-related outcomes. In both tables, we report estimates with and without sibling fixed effects to control for family-level selection. When we include sibling fixed effect, we essentially compare outcomes of first-cousins.

We find robust intergenerational effects of paternal marital distance on children's socioeconomic outcomes. A 10-unit increase in paternal marital distance is associated with a 5.3-percentage-point decrease in the probability that children live on a farm, a comparable 5.5-percentage-point increase in the probability that they reside in an urban area, roughly a 1 point rise in their occupational income score, and a 27% percent increase in the population size of their place of residence. The coefficients with sibling fixed effects are generally smaller in magnitude, but remain economically meaningful and statistically significant. We also find a lower probability of children living in multigenerational families when parental marital distance is high, consistent with greater mobility and a preference for nuclear family living arrangements.

Table 9: The Effect of Marital Distance on Children's Occupational Outcomes

	(1) White-collar	(2) Farmer	(3) Skilled	(4) Unskilled
W/O Sib. FE	.035***	-.049***	.023***	-.008***
	(.0006)	(.0008)	(.0005)	(.0004)
With Sib. FE	.012***	-.018***	.008***	-.002*
	(.0012)	(.0015)	(.0011)	(.0009)
Sample mean	.212	.468	.192	.128
Observations	6,231,900	6,231,900	6,231,900	6,231,900

Note: This table presents the effect of increasing marital distance by 10 units on children's occupational outcomes. The dependent variables are indicators for each of the four occupation categories in Long and Ferrie (2013). The first row ("W/O Sib. FE") displays estimates obtained without parental sibling fixed effects, while the second row ("With Sib. FE") presents estimates obtained with sibling fixed effects. The sample consists of U.S.-born males born after 1750 who were aged 18 or older at the time of the census to which their genealogical records are linked. Additional controls include birth decade × birth state fixed effects and age dummies (in 10-year intervals) at the time of the census. * for p<.05, ** for p<.01, and *** for p<.001.

The positive effect of parental marital distance on children's occupational income is also observed in the types of jobs they hold. Table 9 shows that a 10-unit increase in parental marital distance is associated with a 3.5 percentage-point rise in the probability that a child works in a white-collar occupation and a 5.0 percentage-point decline in the likelihood of farming. Columns (3) and (4) further indicate that greater parental marital distance significantly increases the chance of children holding skilled positions while reducing their likelihood of unskilled employment.

Taken together, these findings show that social integration brought benefits that carried over to the next generation. Children whose parents formed broader marriage ties were more likely to work in white-collar jobs, earn higher incomes, and live in urban, nuclear households—and much less likely to farm or live in multigenerational families. In short, broad marriage networks created lasting social and economic advantages that may have helped drive upward mobility across generations.

7 Conclusion

This paper finds that America's economic transformation fundamentally reshaped its social fabric, causing previously separate communities to become increasingly interconnected. By introducing and analyzing marital distance, a novel metric capturing the network steps separating spouses before marriage, we provide systematic evidence of how industrialization catalyzed social integration across traditional boundaries, creating the "melting pot" effect that became characteristic of American society.

Our findings reveal that average marital distance increased significantly from 10 steps in 1750 to 16 steps by 1950, with the sharpest rise coinciding with America's intensive industrialization period (1870-1915). Counties with greater manufacturing employment consistently experienced larger increases in marital distance, even after accounting for urbanization and demographic shifts. Using a shift-share instrumental variables approach, we establish that this relationship was causal: a ten percentage point increase in manufacturing employment raised marital distance by approximately 1.2 steps. This effect operated through two primary mechanisms: three-quarters stemmed from compositional changes as migration fragmented family networks, while one-quarter reflected actual shifts in preferences toward marrying outside immediate social circles.

The economic consequences of this integration were substantial. Individuals who married at greater social distances experienced improved occupational outcomes, increased geographic mobility, and greater economic independence. These benefits persisted across generations, with children of socially integrated marriages more likely to work in higher-status occupations and urban settings. This pattern suggests that expanding social networks facilitated economic mobility by weakening kinship constraints and broadening access to diverse labor markets.

By systematically measuring social connectedness over two centuries, our results provide new empirical evidence for how economic modernization reshaped social relationships. As America industrialized, it not only changed what people produced but fundamentally altered who they married, where they lived, and how they formed connections across previously separate communities. This transformation weakened traditional boundaries, replacing closed, localized networks with more open, interconnected social systems that became the foundation for American social fluidity and economic dynamism.

References

Abramitzky, R., L. Boustan, and K. Eriksson (2014). A nation of immigrants: Assimilation and economic outcomes in the age of mass migration. *Journal of Political Economy 122*(3), 467–506.

Alesina, A. and P. Giuliano (2014). Family ties. In *Handbook of economic growth*, Volume 2, pp. 177–215. Elsevier.

Bartik, T. J. (1991). *Who Benefits from State and Local Economic Development Policies?* Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Blanc, G. (2024). The Cultural Origins of the Demographic Transition in France. Technical report, Arthur Lewis Lab, The University of Manchester.

Borusyak, K., P. Hull, and X. Jaravel (2022, January). Quasi-Experimental Shift-Share Research Designs. *The Review of Economic Studies 89*(1), 181–213.

Chetty, R., N. Hendren, P. Kline, and E. 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, R., M. O. Jackson, T. Kuchler, J. Stroebel, N. Hendren, R. B. Fluegge, S. Gong, F. Gonzalez, A. Grondin, M. Jacob, D. Johnston, M. Koenen, E. Laguna-Muggenburg, F. Mudekereza, T. Rutter, N. Thor, W. Townsend, R. Zhang, M. Bailey, P. Barberá, M. Bhole, and N. Wernerfelt (2022, August). Social capital I: measurement and associations with economic mobility. *Nature 608*(7921), 108–121.

Enke, B. (2019). Kinship, cooperation, and the evolution of moral systems. *The Quarterly Journal of Economics 134*(2), 953–1019. Publisher: Oxford University Press.

Enke, B. (2020, October). Moral Values and Voting. *Journal of Political Economy 128*(10), 3679–3729. Publisher: The University of Chicago Press.

Fryer, R. G. and S. D. Levitt (2004, August). The Causes and Consequences of Distinctively Black Names. *The Quarterly Journal of Economics 119*(3), 767–805.

Gelfand, M. J., L. H. Nishii, and J. L. Raver (2006). On the nature and importance of cultural tightness-looseness. *Journal of Applied Psychology 91*(6), 1225–1244.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020, August). Bartik Instruments: What, When, Why, and How. *American Economic Review 110*(8), 2586–2624.

Gordon, M. M. (1964). *Assimilation in American life: The role of race, religion, and national origins*. Oxford University Press.

Graham, J., B. A. Nosek, J. Haidt, R. Iyer, S. Koleva, and P. H. Ditto (2011). Mapping the moral domain. *Journal of Personality and Social Psychology 101*(2), 366–385.

Greif, A. (2006, April). Family Structure, Institutions, and Growth: The Origins and Implications of Western Corporations. *American Economic Review 96*(2), 308–312.

Haines, M. R. and Inter-university Consortium for Political and Social Research (2010). Historical, Demographic, Economic, and Social Data: The United States, 1790-2002.

Hornbeck, R., S. Hsu, A. Humlum, and M. Rotemberg (2024, April). Gaining Steam: Incumbent Lock-in and Entrant Leapfrogging.

Hwang, S., D. Jakob, and M. Squires (2024). Health effects of cousin marriage: Evidence from US genealogical records. *AER Insights*.

Hwang, S. I. M. and M. Squires (2024). Linked samples and measurement error in historical US census data. *Explorations in Economic History*, 101579. Publisher: Elsevier.

Kalmijn, M. (1998). Intermarriage and homogamy: Causes, patterns, trends. *Annual review of sociology 24*(1), 395–421. Publisher: Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA.

Long, J. and J. Ferrie (2013). Intergenerational occupational mobility in Great Britain and the United States since 1850. *American Economic Review 103*(4), 1109–1137. Publisher: American Economic Association.

Manson, S., J. Schroeder, D. Van Riper, T. Kugler, and S. Ruggles (2020). National Historical Geographic Information System: Version 15.0.

Mohnenab, M. and M. Viarengocde (2023). America's melting pot: Lessons from the Age of Mass Migration. *Nation Building Big Lessons from Successes and Failures*.

Nelson, M. A. (2020, August). The decline of patrilineal kin propinquity in the United States, 1790–1940. *Demographic Research 43*, 501–532.

Polanyi, K. (1944). *The Great Transformation*. Beacon paperbacks. Farrar & Rinehart, Incorporated.

Price, J., K. Buckles, J. Van Leeuwen, and I. Riley (2021). Combining family history and machine learning to link historical records: The Census Tree data set. *Explorations in Economic History 80*, 101391. Publisher: Elsevier.

Raz, I. T. (2025, February). Soil Heterogeneity, Social Learning, and the Formation of Close-knit Communities. *The Journal of Political Economy*.

Ruggles, S., S. Flood, R. Goeken, J. Grover, E. Meyer, J. Pacas, and M. Sobek (2020). IPUMS USA: Version 10.0.

Triandis, H. C. (1995). New directions in social psychology: Individualism and collectivism.

Weber, M. (1951). *The Religion of China: Confucianism and Taoism*, Volume 93445. Free Press.

Zangwill, I. (1921). *The Melting-Pot*. New York: The American Jewish Book Company.

A Appendix: Supplementary Tables and Figures

Table A.1: Comparison of Marital Distances Across Time Periods (1750-1950)

	(1)	(2)	(3)	(4)
	1750-1799	1800-1849	1850-1899	1900-1949
Random (county)	2.239***	2.495***	2.282***	1.687***
	(0.013)	(0.006)	(0.005)	(0.006)
Random (national)	5.431***	5.407***	4.902***	0.000
	(0.019)	(0.017)	(0.021)	(.)
Observations	395231	1449096	3449900	2180361
Adjusted R2	0.193	0.139	0.087	0.039

Standard errors in parentheses. All specifications include year fixed effects.

Base category is true (observed) marital distance.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure A.1: Marital distance: Path length 6

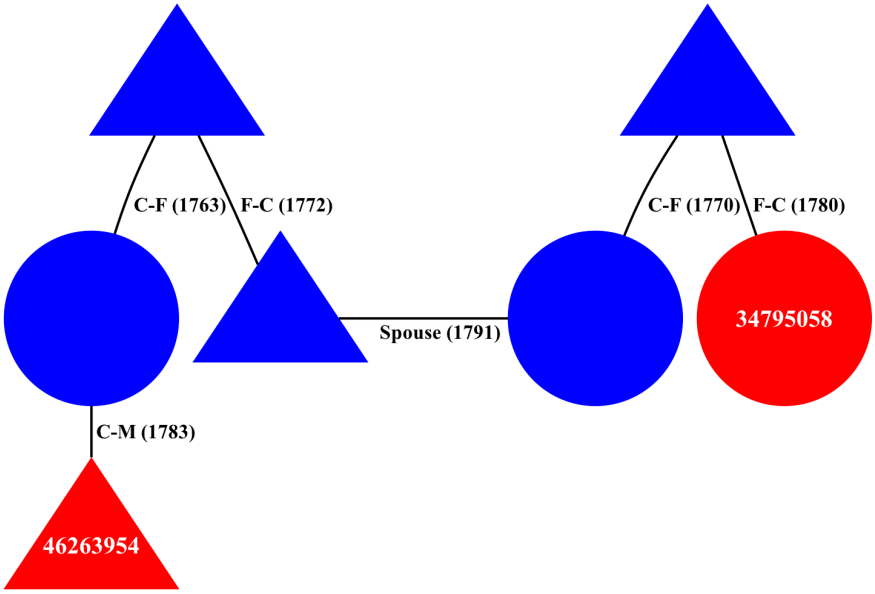


Figure A.2: Marital distance: Path length 12

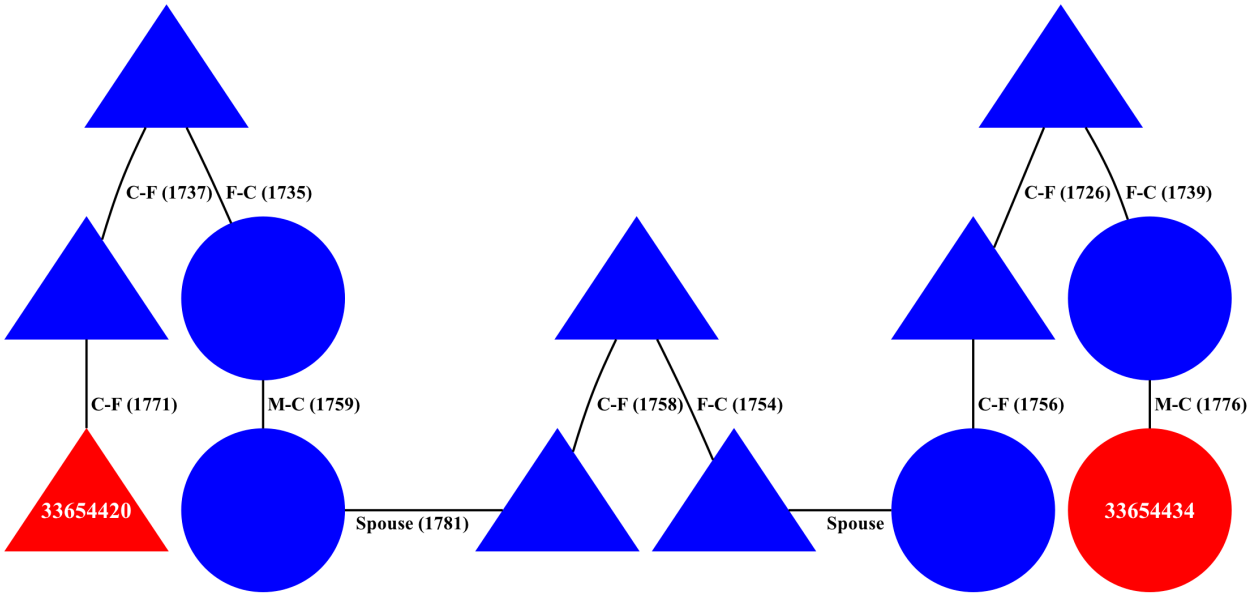


Figure A.3: Marital distance: Path length 12 (vertical)

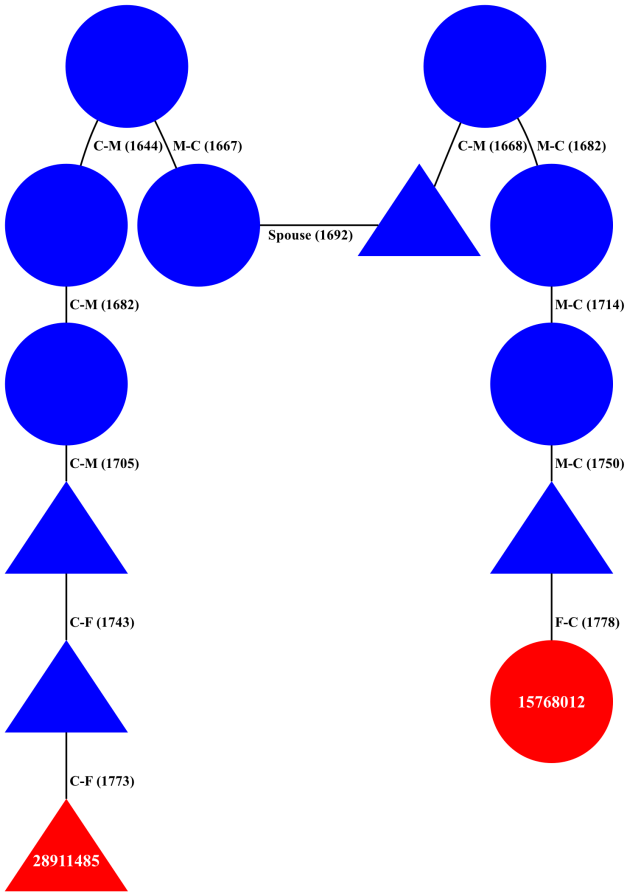


Figure A.4: Marital distance: Path length 15 (vertical)

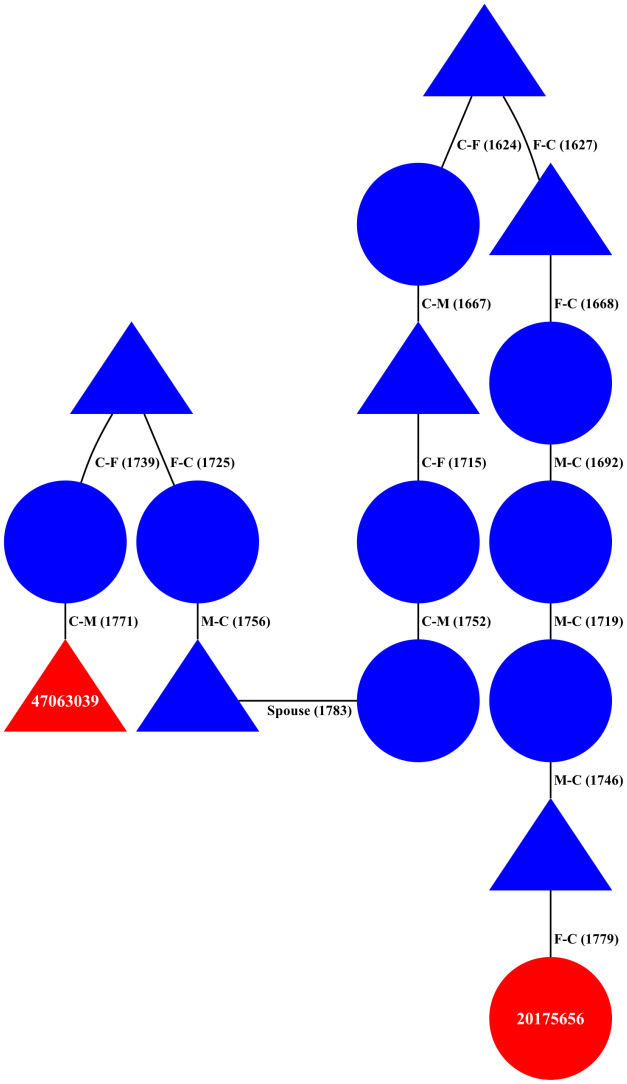


Figure A.5: Marital distance: Path length 16 (horizontal)

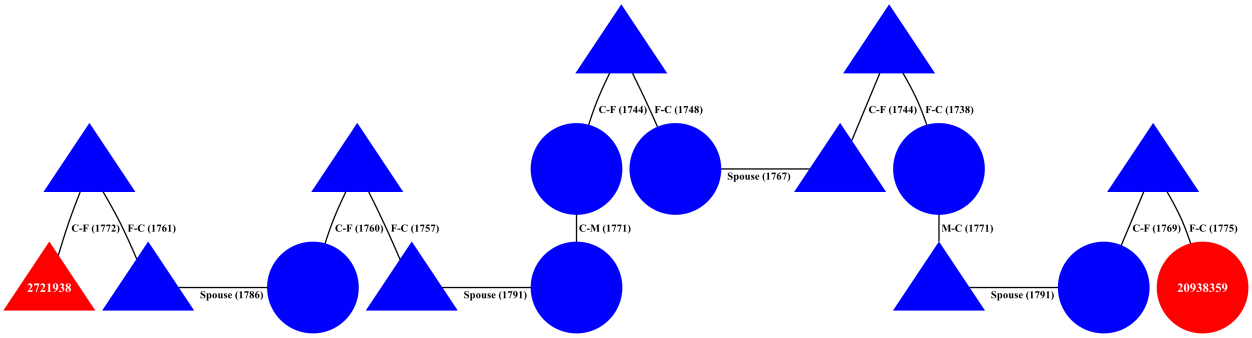


Figure A.6: Marital distance: Path length 24

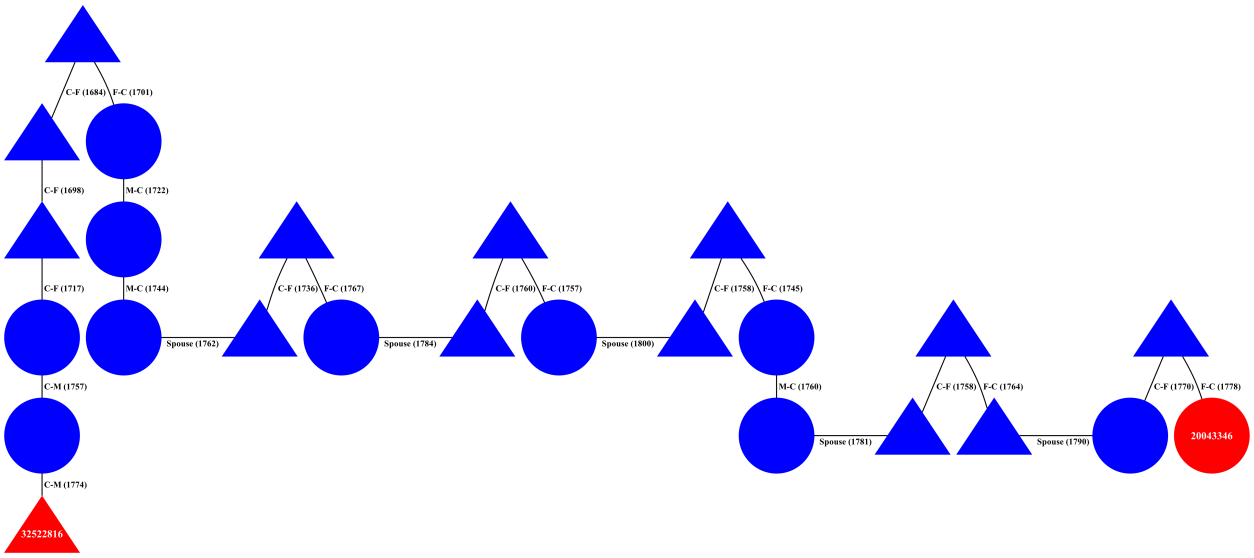


Table A.2: Shift Share: All sample

	(1)	(2)	(3)	(4)	(5)	(6)
Manuf emp share (county)	5.947***	6.214***	2.400***	8.570***	8.716***	3.618***
	(0.926)	(1.125)	(0.293)	(1.352)	(1.379)	(0.559)
Log county population		-0.351	-0.0311		-0.384	-0.0533
		(0.415)	(0.0968)		(0.363)	(0.0804)
Avg. random marital distance			0.687***			0.672***
			(0.0385)			(0.0337)
Observations	16,669	16,669	6,923	16,614	16,614	6,616
Number of counties	3,054	3,054	1,991	2,999	2,999	1,684
Adjusted R2	0.879	0.880	0.965	0.456	0.464	0.823
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
IV (Shift-share)	No	No	No	Yes	Yes	Yes
First stage F-stat				56.81	56.37	36.45

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.3: Shift Share: Excluding Foreign-Borns

	(1)	(2)	(3)	(4)	(5)	(6)
Manuf emp share (county)	5.701***	5.941***	2.350***	8.285***	8.426***	3.575***
	(0.832)	(1.039)	(0.298)	(1.277)	(1.323)	(0.605)
Log county population		-0.299	-0.0348		-0.335	-0.0573
		(0.384)	(0.0993)		(0.336)	(0.0829)
Avg. random marital distance			0.678***			0.662***
			(0.0396)			(0.0344)
Observations	16,635	16,635	6,923	16,580	16,580	6,616
Number of counties	3,052	3,052	1,991	2,997	2,997	1,684
Adjusted R2	0.878	0.879	0.962	0.460	0.466	0.809
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
IV (Shift-share)	No	No	No	Yes	Yes	Yes
First stage F-stat				53.72	53.33	34.42

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.4: Shift Share: ITT (birth county)

	(1)	(2)	(3)	(4)	(5)	(6)
Manuf emp share (county)	-0.00429	0.0341	0.0960	0.0717	0.0956	0.0413
	(0.0982)	(0.0984)	(0.150)	(0.209)	(0.211)	(0.282)
Log county population		-0.0523***	-0.0524*		-0.0533***	-0.0516*
		(0.0151)	(0.0235)		(0.0141)	(0.0201)
Avg. random marital distance			-0.00508			-0.00444
			(0.00753)			(0.00721)
Observations	16,441	16,441	6,923	16,358	16,358	6,616
Number of counties	3,017	3,017	1,991	2,934	2,934	1,684
Adjusted R2	0.893	0.893	0.919	0.908	0.908	0.928
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
IV (Shift-share)	No	No	No	Yes	Yes	Yes
First stage F-stat				56.81	56.37	36.45

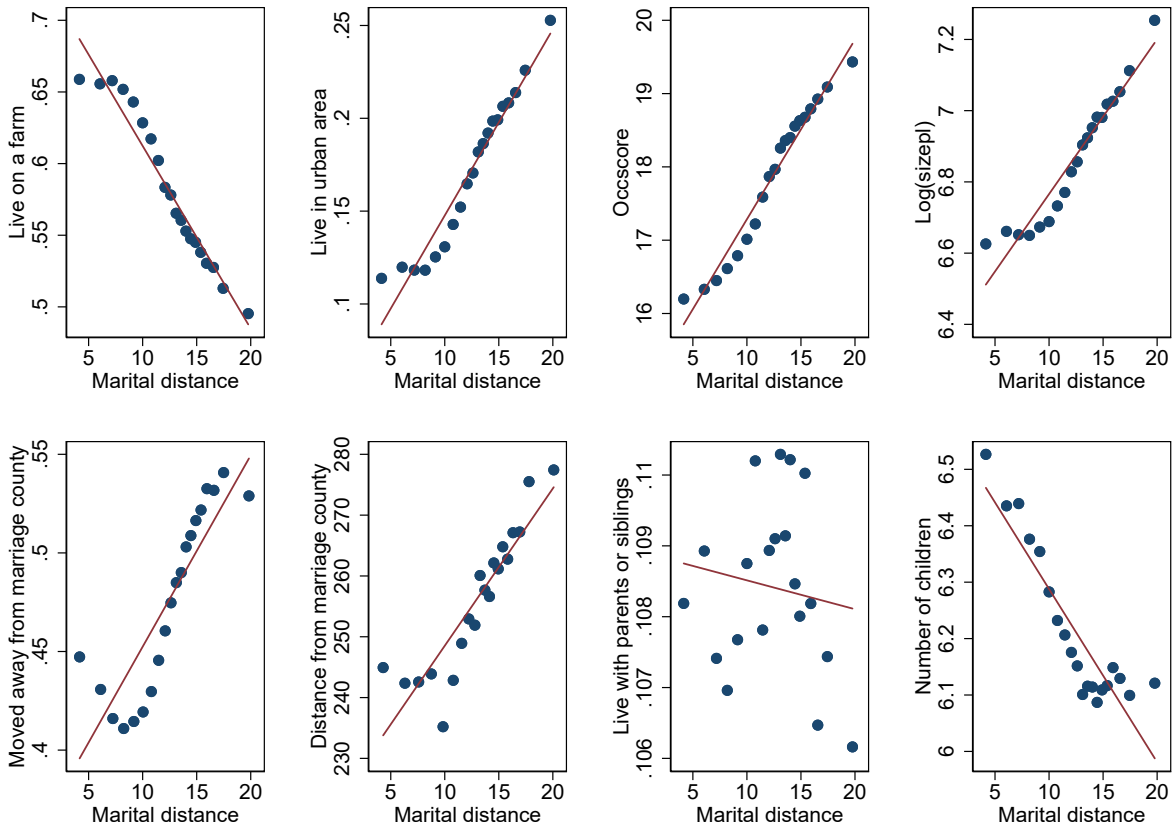
Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Shift Share: birth county - birth decade

	(1)	(2)	(3)	(4)	(5)	(6)
Manuf emp share (county)	-0.237	-0.249	-0.285	-0.572	-0.595	-0.887*
	(0.142)	(0.142)	(0.184)	(0.315)	(0.319)	(0.372)
Log county population		0.0198	0.0215		0.0231	0.0306
		(0.0183)	(0.0235)		(0.0167)	(0.0213)
Avg. random marital distance			-0.00901			-0.00232
			(0.00754)			(0.00727)
Observations	15,860	15,860	6,920	15,783	15,783	6,613
Number of counties	3,017	3,017	1,990	2,940	2,940	1,683
Adjusted R2	0.024	0.024	0.051	0.001	0.001	0.003
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
IV (Shift-share)	No	No	No	Yes	Yes	Yes
First stage F-stat				45.61	45.56	28.68

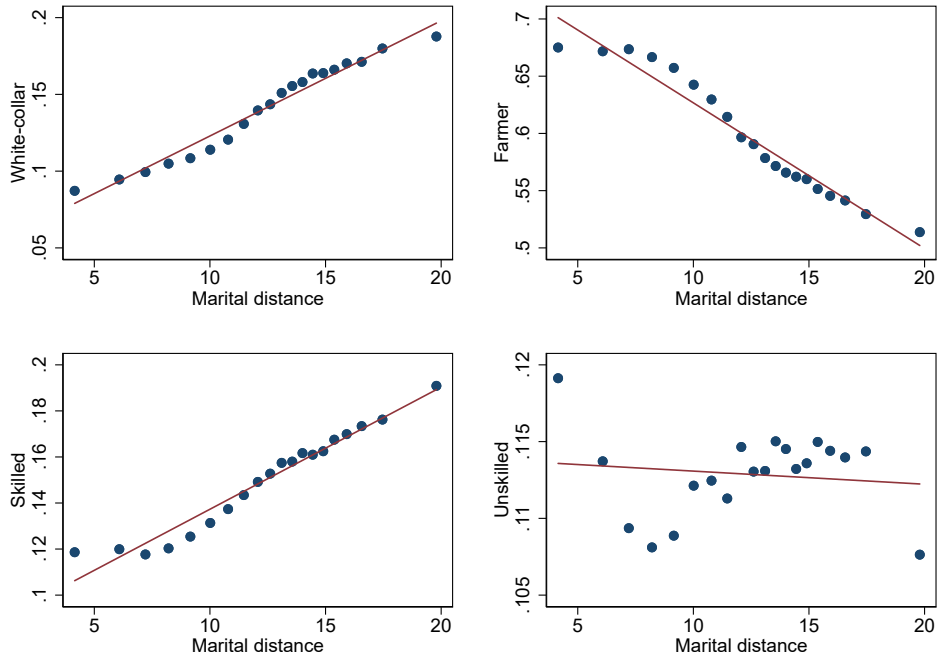
Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure A.7: Binsreg



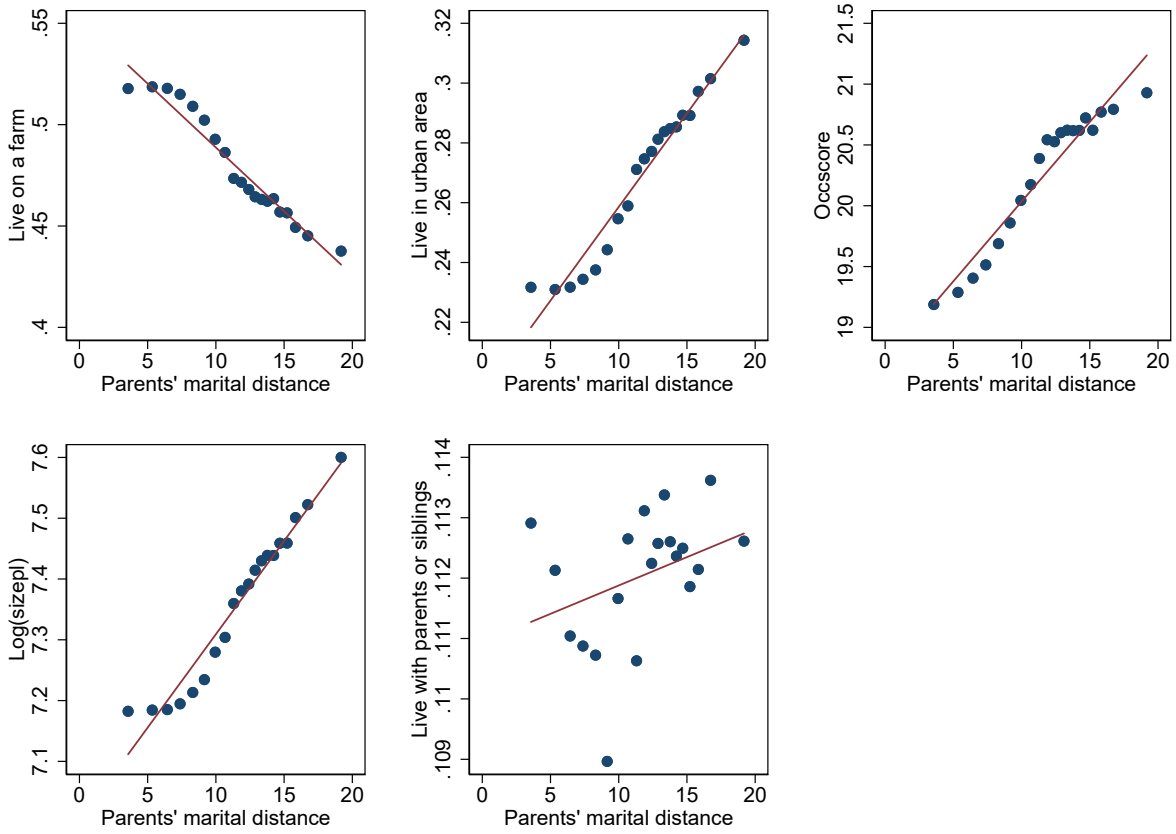
Note: birth decade X birth state dummies and dummies for age in 10 years at census dummies controlled for. sample: male, US born, birth year $\geq$ 1750 and marriage year $\geq$ 1800

Figure A.8: Binsreg: occupation variables



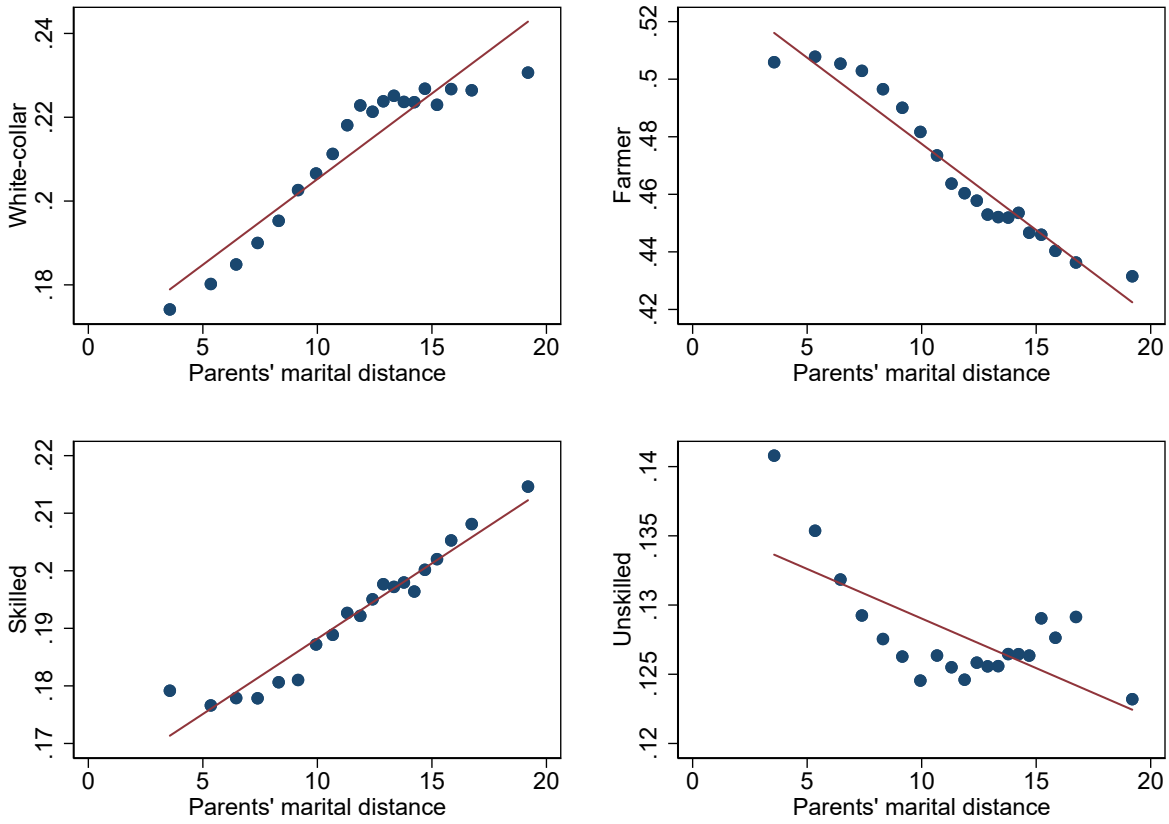
Note:

Figure A.9: Binsreg: next generation's outcome



Note: birth decade X birth state dummies and dummies for age in 10 years at census dummies controlled for. sample: male, US born, birth year $\geq$ 1750 and above age 18 at the time of census

Figure A.10: Binsreg: next generation's occupational outcome



Note: birth decade X birth state dummies and dummies for age in 10 years at census dummies controlled for. sample: male, US born, birth year $\geq$ 1750 and above age 18 at the time of census

Table A.6: The Effect of Marital Distance on Socio-Economic Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Farm	Urban	Occscore	Log(sizepl)	Moved to Diff. County	Distance from Marriage county (in miles)	Live with parents(-in-law)	Number of Children
W/O Sib. FE	-.116***	.094***	2.255***	.409***	.086***	22.026***	-.001***	-.262***
	(.0007)	(.0006)	(.0168)	(.0025)	(.001)	(1.0635)	(.0003)	(.006)
With Sib. FE	-.042***	.025***	.838***	.1***	.026***	-11.435***	-.007***	.156***
	(.0015)	(.001)	(.0368)	(.0042)	(.0024)	(3.1609)	(.0009)	(.0141)
Sample mean	.582	.171	17.863	6.87	.475	255.865	.108	6.214
Observations	4,338,295	4,338,295	4,338,295	4,338,295	3,384,236	1,592,620	4,338,295	4,292,191

Note: This table presents the effect of increasing the marital distance by 10 units on various socio-economic outcomes. The first row ("W/O Sib. FE") displays estimates obtained without sibling fixed effects, while the second row ("With Sib. FE") presents estimates obtained with sibling fixed effects. The sample for this table consists of U.S.-born males born after 1750 who were aged 18 or older at the time of the census to which their genealogical records are linked. Additional controls include birth decade × birth state fixed effects and age dummies (in 10-year intervals) at the time of the census. * for p<.05, ** for p<.01, and *** for p<.001.

Table A.7: The Effect of Marital Distance on Occupational Outcomes

	(1) White-collar	(2) Farmer	(3) Skilled	(4) Unskilled
W/O Sib. FE	.069***	-.115***	.049***	-.003***
	(.0005)	(.0008)	(.0005)	(.0004)
With Sib. FE	.025***	-.044***	.014***	.004***
	(.0011)	(.0016)	(.0011)	(.001)
Sample mean	.141	.597	.15	.113
Observations	3,939,975	3,939,975	3,939,975	3,939,975

Note: This table presents the effect of increasing marital distance by 10 units on occupational outcomes. The dependent variables are indicators for each of the four occupation categories in Long and Ferrie (2013). The first row ("W/O Sib. FE") displays estimates obtained without sibling fixed effects, while the second row ("With Sib. FE") presents estimates obtained with sibling fixed effects. The sample consists of U.S.-born males born after 1750 who were aged 18 or older at the time of the census to which their genealogical records are linked. Additional controls include birth decade × birth state fixed effects and age dummies (in 10-year intervals) at the time of the census. * for p<.05, ** for p<.01, and *** for p<.001.