

From Roots to Branches^{*}

The Role of Families in Structural Change

Luca Looser[†]

October 2026

Updated regularly, please click [here](#) for the latest version.

Abstract

Economic opportunities differ sharply across sectors and places, but workers face barriers to reaching them. Kinship is often viewed as one such barrier because relatives can anchor workers to their current sector and location. Yet kin in expanding sectors and at destinations can facilitate sector switching and migration by creating such connections for other relatives. This paper studies how these opposing forces shape U.S. structural transformation from 1880 to 1970. I reconstruct extended kin for 32 million individuals and use deaths of relatives to show that sector-employed kin increase same-sector employment and local kin reduce migration, while migrants disproportionately move toward destinations where relatives already live. I embed these forces in a dynamic spatial model in which workers' choices affect other kin and later generations. Kin account for roughly 10–15 percent of the sectoral and migration barriers. Their initial concentration in agriculture and less industrialized places slows structural transformation, but evolving kin connections offset about 90% of this drag by 1970, raising the manufacturing share by 2.05 percentage points (2.6%) and advancing the transition by about seven years. With a moving voucher, ignoring spillovers to other kin understates long-run manufacturing gains by 26% and earnings gains by 15%, while the same budget induces about three times as many moves when more kin live at eligible destinations.

^{*}I am grateful to Elisa Giannone and Dávid Krisztián Nagy for continuous support. I thank Francesco Saverio Lenzi, Dehua Li and Tancredi Rapone for discussing this paper. This paper has also benefited from discussions with Alexander Bick, Francisco Buera, Alessandro Cerusico, Bruno Conte, Marta Cota, Diego Daruich, Kevin Donovan, Maximiliano Dvorkin, Fabian Eckert, Jan Eeckhout, Manuel Garcia-Santana, Madalena Gaspar, Ed Glaeser, Albrecht Glitz, Philipp Grübener, Marko Irisarri, Adam Jørring, Tim Kehoe, Tatjana Kleineberg, Gianmarco León-Ciliotta, Remy Levin, Miquel Lorente, Niclas Moneke, Joan Monras, Robert Pollak, Giacomo Ponzetto, Tommaso Porzio, Nick Pretzner, Hugo Reichardt, Hannah Rubinton, Edouard Schaal, Caterina Soto Vieira, Jaume Ventura, Daniela Vidart and Tuomo Virkola. I thank participants at the CREi International Lunch, UPF Internal Applied Lunch, BSE Jamboree 2025, 20th End-of-Year Conference of Swiss Economists Abroad, QMUL PhD Workshop, EUI ECO PhD Workshop, the HEC PhD Workshop, the XXIX Workshop on Dynamic Macroeconomics, WashU Macro Breakfast, the LAEF/UConn Conference, the 20th North American Meeting of the Urban Economics Association and the OxDev 2026 Oxford Development Economics Workshop for their feedback. Research reported in this publication was supported by UPF Scientific Computing Core Facility. I am particularly grateful to Francisco Buera for hosting me at Washington University in St. Louis and to Hannah Rubinton for hosting me at the Federal Reserve Bank of St. Louis, where part of this project was completed. All errors are my own.

[†]Department of Economics, Universitat Pompeu Fabra, Carrer de Ramon Trias Fargas, 25–27, Sant Martí, 08005 Barcelona. Email: luca.looser@upf.edu.

1 Introduction

Structural transformation—the movement of workers out of agriculture toward manufacturing and services—is a central feature of economic growth (Kuznets, 1973). Because productivity differs across sectors and locations, barriers to these movements can slow structural change and reduce aggregate output. Kinship and family connections in particular are often viewed as one such barrier because relatives can keep workers in their current sector and location (Alesina & Giuliano, 2014; Enke, 2019; Ghosh, Hwang, & Squires, 2023; Henrich, 2020; Schulz, 2022; Schulz, Bahrami-Rad, Beauchamp, & Henrich, 2019). However, kin also help workers reach new opportunities: relatives in expanding sectors help them find jobs, while kin at destinations make migration easier. As workers themselves switch sectors and migrate, they create these connections for other relatives and later generations. Yet little is known about how these two sides of kinship interact over development. How much do origin and same-sector kin slow movement into new sectors and places? How much do relatives in possible destinations and expanding sectors help workers make these transitions? How do these effects build across generations and what do they imply for structural transformation and spatial policies?

I study these questions during U.S. structural transformation from 1880 to 1970 and make four contributions. First, I construct new data that trace extended kin across sectors, locations and generations and document how the distribution of relatives changes during structural transformation. Second, I provide empirical evidence on both sides of the mechanism: kin in current sectors and origins reinforce persistence, while kin in alternative sectors and at destinations facilitate sector switching and migration. Third, I develop a dynamic spatial model to show how these forces change the pace and level of structural transformation. Fourth, I use a large-scale moving subsidy to study whether temporary spatial policies, such as moving vouchers, can have lasting effects beyond its direct recipients across kin and generations.

Studying this mechanism requires large-scale data that link individuals to extended kin and observe where those relatives live and work across generations. Kinship is particularly useful for this purpose because family relationships are largely given by birth rather than formed through common workplaces or locations and can remain important after relatives move apart and be measured across long time spans (McPherson, Smith-Lovin, & Cook, 2001; Roberts & Dunbar, 2011). I construct such data by combining the full-count U.S. censuses with Census Tree links and recursively exploiting household relationships to recover extended relatives across generations (Buckles, Haws, Price, & Wilbert, 2025; Ruggles et al., 2024). This construction reaches relatives who need not themselves appear in the ancestry links and retains their individual census information, including sector and location. By 1940, the resulting data contain approximately 32 million individuals whose kin histories can be traced back to 1850.

The data first show that extended kin move with structural transformation beyond the vertical parent-children links. Among aunts and uncles, for example, the manufacturing share rises from

about 10% in 1880 to roughly 40% by 1930, while the mean distance among cross-CZ connections rises from 542 to 849 kilometers and the share extending more than 1,000 kilometers nearly doubles. I then use reduced-form evidence to show that where relatives live and work affects individual choices. Early deaths of sector-employed relatives reduce subsequent employment in the deceased relative's sector, while deaths of local non-employed relatives increase migration. I also show that, conditional on origin and common origin–destination attractiveness, migrants disproportionately choose destinations where relatives are already established.

I bring these empirical forces together in a novel dynamic spatial model in which workers choose sectors and locations and relatives in the current sector and location can discourage switching and migration, while relatives in other sectors and at destinations can facilitate them. Workers' choices then change these connections for others: entering manufacturing gives other kin another relative in manufacturing, while migrating gives them a relative at a new destination. Workers do not take into account how these choices affect other relatives beyond their own descendants, creating spillovers across kin and generations. These externalities can reinforce the concentration of relatives in particular sectors and places, as one worker's choice makes the same choice more attractive to other kin. Whether this feedback slows or accelerates structural transformation depends on where kin are located and their strength: connections to agriculture and origins reinforce persistence, while connections created in manufacturing and new destinations can propagate the transition across generations.

I quantify the model for the United States from 1880 to 1970. Deaths of parents and extended kin identify how strongly kin affect sector and location choices. Using indirect inference, an agricultural productivity shock then provides external variation for estimating how strongly workers respond to changing economic opportunities, informing key choice elasticities in the framework.

The quantitative analysis shows that kinship is a sizable part of the barriers workers face when changing sectors and places: in 1910, removing kin connections requires adjustments equal to roughly 10–15 percent of the sectoral and migration barriers. Their aggregate effect, however, depends on where relatives live and work. In 1880, kin are concentrated in agriculture and in less industrialized places, so this distribution slows structural transformation. By 1950, their presence lowers the manufacturing share by 1.31 percentage points, or 1.8 percent, relative to an economy in which kin do not affect sector and location choices. Almost all of this drag comes from where relatives are located. As workers enter manufacturing and migrate, however, they place more relatives in manufacturing and in new destinations. These changes offset 55 percent of the initial drag by 1950 and about 90 percent by 1970, raising the 1970 manufacturing share by 2.05 percentage points and advancing the transition by about seven years. Because these gains are concentrated in initially less industrialized regions, evolving kin connections also help spread industrialization to places that begin the transition further behind ([Eckert & Peters](#),

2022).

These spillovers matter for policies that seek to move workers out of lagging regions. Relocation assistance has been proposed as a way to help workers reach better opportunities and large-scale programs have permanently relocated households in developing economies, including China's Anti-Poverty Relocation Program and Indonesia's Transmigration Program (Bazzi, Gaduh, Rothenberg, & Wong, 2016; Zhang, Xie, & Zheng, 2023). I study a one-generational moving voucher targeted at workers in low-manufacturing places. Because a subsidized move places a relative at a new destination, the policy can affect kin who never receive the voucher and later generations, an effect the initial mover does not take into account. Ignoring this kinship spillover understates the voucher's long-run effect on manufacturing by 26 percent, its earnings effect by 15 percent and the earnings-based welfare gain of later generations by about 12 percent. It also misses an important targeting margin: among economically attractive destinations, the same budget induces about three times as many moves when more kin already live there. Evaluating relocation policy only through its direct recipients therefore understates its long-run effects and can mismeasure where scarce policy resources generate the largest impact.

Related literature This paper first contributes to work on barriers to structural transformation and labor reallocation. Beyond productivity and demand (Boppart, 2014; Comin, Lashkari, & Mestieri, 2021; Ngai & Pissarides, 2007), the pace of structural change depends on workers' ability to move across sectors and places. Existing work emphasizes sector-specific human capital and experience, worker heterogeneity and costly sector switching (Artuç, Chaudhuri, & McLaren, 2010; Dix-Carneiro, 2014; Herrendorf & Schoellman, 2018; Lee & Wolpin, 2006; Porzio, Rossi, & Santangelo, 2022), while spatial models study how sectoral reallocation interacts with migration frictions, local productivity and trade (Adamopoulos, Brandt, Chen, Restuccia, & Wei, 2024; Budí-Ors & Pijoan-Mas, 2025; Cao, Chen, Xi, & Zuo, 2024; Eckert & Peters, 2022).¹ Kinship and local social ties are one source of these barriers: origin-based insurance and attachment can inhibit rural-urban, off-farm and geographic mobility (Ghosh et al., 2023; Munshi & Rosenzweig, 2016; Zabek, 2024). I study this source of reallocation frictions and show how its effect changes endogenously with development. Past reallocation changes where relatives live and work and therefore the sectors and places that later generations can access, providing a microfoundation for why barriers to structural change can weaken, strengthen or shift as development proceeds.

The paper is most directly related to work on kin and social connections in migration and labor-market allocation. In migration, local connections can discourage leaving while connections at destinations facilitate moves and shape where migrants go (Blumenstock, Chi, & Tan, 2025; Büchel, Ehrlich, Puga, & Viladecans-Marsal, 2020; Ilyés, Boza, Lőrincz, & Eriksson, 2023;

¹See Donovan and Schoellman (2023) for a review of labor-market frictions in structural transformation.

Munshi, 2003; Stuart & Taylor, 2021).² Quantitative work embeds social connections in spatial allocation (Koenen & Johnston, 2024; Sahai & Bailey, 2023; Zabek, 2024; Zerecero, 2021). A parallel literature shows that family and social connections shape access to employers, occupations and intergenerational specialization (Cassan, Keniston, & Kleineberg, 2021; Kramarz & Nordström Skans, 2014; Lo Bello & Morchio, 2022; Munshi, 2011; Patel & Vella, 2013), while stronger kinship can generate both geographic and occupational persistence (Ghosh et al., 2023).³ Existing work therefore points to an origin–destination distinction: connections to current places or activities can create attachment, while connections to alternatives can provide access. I bring these margins together within the same kin structure and across both sector and space. By tracing extended kin across sectors, places and generations, the data allow me to disentangle these forces and quantify their distinct roles in structural transformation. A further step is to allow these connections themselves to evolve as workers change sectors and locations. This links the paper to models in which social interactions shape occupational or spatial allocation (Cassan et al., 2021; Eckert & Kleineberg, 2024; Koenen & Johnston, 2024; Sahai & Bailey, 2023) and respond endogenously to economic choices (Fan & Zou, 2021; Kim, Patacchini, Picard, & Zenou, 2023). Closest to my analysis, Dai, Mookherjee, Munshi, and Zhang (2020) show that hometown business connections generate sectoral and spatial concentration, while Dai, Mookherjee, Munshi, and Zhang (2023) document that connections accumulated during an earlier transition can facilitate entry into a new activity while impeding subsequent reallocation. In my setting, sectoral and spatial reallocation reshape the distribution of kin across sectors and locations, changing both attachment to inherited choices and connections to alternative opportunities for subsequent generations. This provides a kinship-based microfoundation for persistence and path dependence (Allen & Donaldson, 2020).

Finally, the paper contributes to work using linked historical and genealogical data. Population-scale genealogies reconstruct extended kinship over long horizons (Kaplanis et al., 2018; Koylu, Guo, Huang, Kasakoff, & Grieve, 2021), while census-linking projects follow individuals and families across generations (Althoff, Brookes Gray, & Reichardt, 2026; Buckles et al., 2025; Helgertz et al., 2022). Existing work primarily uses these records to study intergenerational mobility and persistence (Feigenbaum, 2018; Long & Ferrie, 2013; Ward, 2023).⁴ I extend these approaches by using public available ancestry links as a starting point and recursively exploiting full-count census household relationships to recover additional extended relatives who need not themselves appear in the ancestry tree data (Espín-Sánchez, Ferrie, & Vickers, 2023). Because these relatives remain linked to their census records, I observe where they live and which sectors

²Origin-based insurance and temporary migration are studied by Munshi and Rosenzweig (2016) and Morten (2019); see Munshi (2020) for a review.

³Related evidence on labor-market connections and social-ties support includes Glitz (2017), Compton and Pollak (2014) and Bau and Fernández (2021).

⁴For multigenerational and extended-family persistence, see Lindahl, Palme, Sandgren Massih, and Sjögren (2015) and Adermon, Lindahl, and Palme (2021).

they work in. This makes it possible to reconstruct extended kin across generations together with their economic and geographic characteristics and to trace how these connections evolve during U.S. structural transformation.

The remainder of the paper proceeds as follows. Section 2 describes the construction of family networks. Section 3 presents the empirical evidence. Section 4 develops the quantitative model and Section 5 describes its estimation. Section 6 reports the counterfactual exercises. Section 7 concludes.

2 Constructing family trees

I construct a new census-scale panel of extended family connections that follows relatives across households, generations and locations from 1850 to 1940. The key difference is to recover these connections directly from census records rather than relying on pre-existing genealogies. Existing population-scale genealogies take family relationships from pre-constructed family relationships (Kaplanis et al., 2018; Koylu et al., 2021), while census-based work often reconstructs only specific multigenerational links (Espín-Sánchez et al., 2023). I instead combine full-count U.S. census microdata with cross-census individual links from the Census Tree Project (Buckles et al., 2025) and recursively use within-household relationships to recover extended kin.⁵⁶ A relative can therefore enter the family structure through a linked household member even without being linked across census waves directly, broadening coverage. As households split, new generations appear and relatives move, the procedure carries family identifiers forward while retaining the economic and demographic information recorded in each census.

Constructing family trees and relationships. The construction combines cross-census links, which follow individuals over time, with household relationship codes, which identify relatives observed around them in each census. I start in 1850 by assigning household heads and their observed blood and half-blood relatives to common family lineages using IPUMS relationship-to-head codes.⁷⁸ I then follow these individuals into later censuses using the Census Tree Project crosswalks. When a linked person appears in a new household, newly observed relatives are added to the existing lineage and can carry it forward in subsequent waves. Repeating this process through 1940 produces extended family trees spanning households, locations

⁵⁶The main empirical sample uses family trees from 1850 to 1940. Because Census Tree Project crosswalks are unavailable for 1950, extending the construction to 1950 substantially reduces coverage.

⁶Alternative linking procedures were explored. The Census Tree Project performs well in terms of scope and representativeness; see Althoff, Gray, and Reichardt (2025) for a systematic comparison. The crosswalks constructed by Althoff et al. (2025) improve accuracy along some dimensions but yield a smaller linked sample.

⁷⁸This excludes unrelated co-residents such as lodgers and domestic workers.

⁸Using household heads as initial anchors creates some dependence on household definitions, but this dependence declines as additional generations are incorporated.

and generations. I use individuals' positions within these trees to recover pairwise relationships. Because an incorrect person link could otherwise spread to many relatives, I extend a lineage only when the implied relationships are consistent with the one observed in the new census. I also restrict propagation through spousal, descendant and collateral links when they would generate ambiguous assignments across lineages. Appendix [A1.1](#) describes these checks and propagation rules as well as the overall algorithm in detail. The resulting family identifiers can therefore follow the same lineage even after its members no longer live together.

I next recover how individuals within each lineage are related. I assign each person a genealogical position in the family tree and compare these positions to identify parents and children, siblings, grandparents, aunts and uncles, nieces and nephews and cousins. Relationships that cannot be uniquely inferred are treated as ambiguous and are not used to extend blood lineages. Spouses do not enter one another's blood lineage, while children inherit identifiers from both parental branches and can therefore belong to multiple lineages. When the same pair is connected through more than one lineage, I retain the relationship implied by the closest common ancestor so that each pair enters a given kin measure only once.⁹ Figure [A1.1.4](#) illustrates an example of the resulting family tree.

Coverage and representativeness. With this approach, roughly 30–40% of the U.S.-born adult population observed in 1940 can be linked to familial ancestors over the preceding 90 years.¹⁰ This yields approximately 32 million individuals in 1940 for whom at least one reconstructed family lineage can be traced back to 1850 (Table 1).

Two potential sources of error arise in the reconstruction. First, within-census household relationships may be miscoded or internally inconsistent. I therefore apply consistency checks before using them to extend family connections. Second, cross-census links may match the wrong individual. I reduce this risk by requiring support from at least two Census Tree linkage sources. These restrictions favor accuracy over coverage but cannot eliminate linkage error entirely.

A separate limitation is that the reconstructed family tree is observed only through individuals who appear in the census data at a given date. I therefore recover the relatives observed in each census wave rather than a permanent genealogy containing all family relationships. This differs from genealogy-based data constructed from pre-existing ancestry trees, such as [Koylu et al. \(2021\)](#), where family links can be retained independently of whether both relatives are observed in a particular census. The advantage of the census-based approach is that relatives can be linked directly to their contemporaneous location and economic outcomes, but the set

⁹Remaining ties are resolved using the reconstruction order and, if necessary, the lower family identifier.

¹⁰These rates are broadly consistent with chaining decade-by-decade linkage success documented in [Althoff et al. \(2025\)](#). To reduce linkage error, I require support from at least two of the linkage sources incorporated in the Census Tree Project.

of observed kin varies with census coverage and linkage.

The construction also generates selection. Linkage success varies systematically with observable characteristics, so the linked sample need not reproduce the contemporaneous full census. In addition, tracing a lineage over several decades requires family members to survive, remain observable in the census and be successfully linked across successive waves, which can further change the composition of the longitudinal sample (Antonie, Inwood, Minns, & Summerfield, 2020; Ruggles, Fitch, & Roberts, 2018). Relative to the full census, the unweighted linked sample is more likely to be white, rural, farm-resident and employed in agriculture, whereas differences in age and sex are considerably smaller (Table A1.4.2).

To reduce these observable compositional differences, I reweight the linked sample toward the corresponding full-count census using inverse-propensity weights following Ward (2023). Separately by census wave, I estimate the probability of appearing in the linked sample as a function of demographic, socioeconomic and geographic characteristics and weight linked individuals by the inverse of this probability.¹¹ The weights balance the linked sample toward the corresponding census population along the observed characteristics included in the propensity model and substantially reduce the differences between the linked and full-count samples (Bailey, Cole, & Massey, 2020). They cannot, however, correct selection on unobservables, restore family connections that are not recovered or eliminate false links.

Table A1.4.1 summarizes family size and network depth. The median reconstructed family contains 22 linked members, with an interquartile range of 9 to 51 and spans roughly three generations from the initial anchor.¹² Figure A1.4.1 shows a fat-tailed family-size distribution, with a small number of large reconstructed lineages arising as distant lateral ties accumulate across waves.¹³ Networks with only 1–5 members mainly reflect linkage attrition and the conservative exclusion of non-relatives and marriage-based upstream connections. Table 1 reports the average number of relatives by relationship type from 1880 to 1940. Extended kin account for large share of the connections. Appendix A1.3 and A1.2 compares the reconstructed geographic family network with independent measures of social and migration connectedness as well as internal validation checks.

Inferring deaths. To obtain family-specific shocks, I infer deaths from census records. Since population-wide individual vital statistics are unavailable before 1950, I combine repeated observations of marital status with longitudinal spouse links to infer mortality. IPUMS records whether individuals are married, single, divorced or widowed from 1880 onward. I classify a spousal death between census years t and $t + 1$ when an individual is married at t , widowed at $t + 1$ and the spouse observed at t cannot be found in the $t + 1$ linkage universe. I then

¹¹See Appendix A1.1.2 for details.

¹²This includes total family members, i.e. second and third cousins for example.

¹³The empirical analysis focuses on ties no more distant than nieces and cousins.

Table 1: Coverage and composition of reconstructed kin networks, 1880–1940

	1880	1900	1910	1920	1930	1940
Panel A. Composition of the kin network						
<i>Vertical kin</i>						
Parent	1.80	1.70	1.44	1.79	1.71	1.74
Child	1.74	1.69	1.42	1.79	1.71	1.74
Grandparent	0.42	0.55	0.32	0.39	0.38	0.33
Grandchild	0.44	0.55	0.33	0.41	0.38	0.35
<i>Horizontal kin</i>						
Sibling	3.48	3.06	2.46	2.99	2.64	2.45
Uncle/Aunt	1.59	1.70	1.42	1.54	1.45	1.38
Niece/Nephew	1.64	1.71	1.43	1.55	1.47	1.40
First cousin	2.39	2.97	2.29	2.57	2.27	2.11
All relatives	13.50	13.93	11.11	13.03	12.01	11.50
Panel B. Sample coverage						
Individuals in network (millions)	17.46	21.96	24.72	27.82	31.41	32.21
<i>Share of full population (%)</i>	34.9	29.0	26.9	26.3	25.6	24.4
<i>Share of U.S.-born population (%)</i>	51.6	43.4	40.3	38.2	35.4	31.2

Notes: Panel A reports the mean number of relatives of each type per individual in the reconstructed kin network, pooling relatives across commuting zones and employment sectors. Relationships are classified from reconstructed family paths. “All relatives” is the sum of the eight reported blood-relationship categories. Panel B reports the number of individuals in the network and its coverage relative to the full and U.S.-born census populations in each year.

propagate the implied death through the reconstructed family network so that relatives can be assigned the shock even when they live in different households or locations.

The inferred-death measure has three main limitations. First, marital-status reporting error can generate misclassification, while deaths followed by remarriage before the next census may not be detected. Second, the procedure identifies deaths only for married individuals whose surviving spouse remains observable and successfully linked across waves, so the resulting death sample is selected. Third, failures to link a surviving spouse can generate false inferred deaths. Table A1.6.1 reports summary statistics for the inferred-death sample. Appendix A1.5 documents correlations with state-level mortality rates. These shocks form the basis of the empirical analysis in the next section.

3 Family ties and reallocation of labor

With the data in hand, this section asks how family connections shape workers’ sector and location choices. Kinship is particularly useful for studying persistence in access across gen-

erations for three reasons. First, family relationships are largely given by birth rather than formed because people work in the same sector or live in the same place, decreasing the risk of selection. Second, these relationships can remain important even after relatives move apart or interact less often (Roberts & Dunbar, 2011). Third, when one family member migrates, that move can make the same destination easier to reach for other relatives (Baseler, 2025). Kinship therefore provides a natural way to study how access created by one person’s choices can persist and affect other family members over time. Importantly, these effects are distinct from broader origin-based connections. Appendix A2.2 shows that family connections predict sector and destination choices beyond migration patterns shared by people from the same hometown or birthplace.

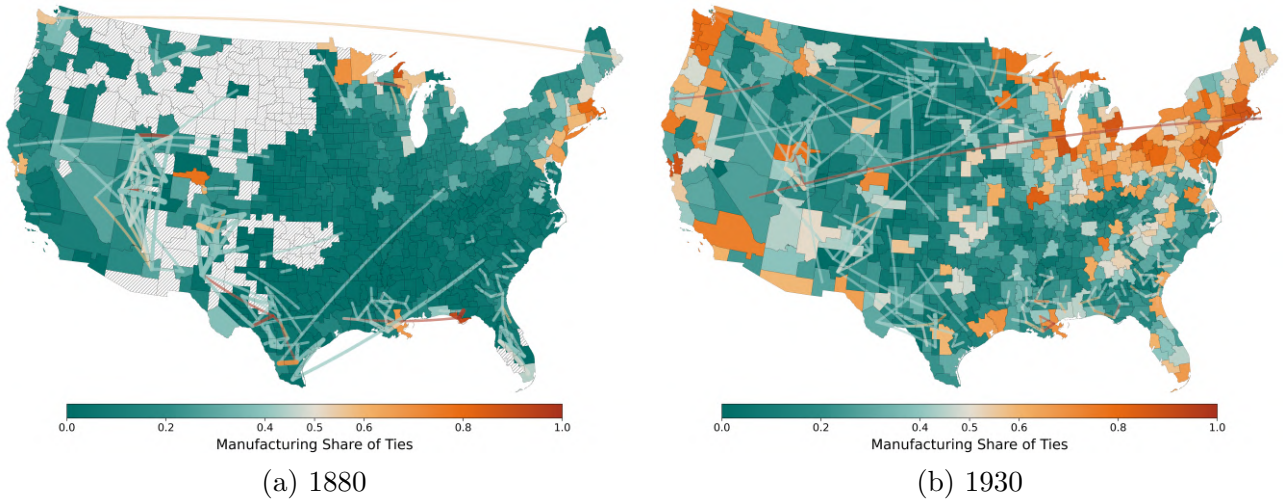
I separate the evidence into two parts. I first document how family connections change over time and how workers’ sector and destination choices covary with where their relatives work and live. These relationships are descriptive because the location and sector of relatives are themselves endogenous. I then use deaths of relatives to generate plausibly exogenous changes in family connections. Deaths of sector-employed relatives generate changes in sector-specific family presence, while deaths of local non-employed kin yield differences in local family access.

3.1 Descriptive evidence

Changing family connections. Family connections change along both margins of structural transformation. As manufacturing expands, relatives increasingly work in manufacturing: among aunts and uncles, for example, the manufacturing share rises from about 10% in 1880 to roughly 40% by 1930, both within and outside the individual’s CZ, with a similar increase among cousins. At the same time, relatives become more dispersed across space. Among cross-CZ ties, mean distance rises from 542 to 849 km and the median from 306 to 475 km, while the share extending more than 1,000 km nearly doubles from 16.5% to 30.1%. Appendix A2.1 reports the relationship-specific patterns and full distance distribution.

Figure 1 shows where these changes occur. The background gives each CZ’s manufacturing share, while the lines highlight strong family connections between CZ pairs. To separate bilateral concentration from the fact that some CZs simply have more family ties overall, I use relative connectedness, $RC_{od} = S_{od}/(S_o \times S_d)$ (Bailey, Cao, Kuchler, Stroebe, & Wong, 2018a). Here S_{od} is the observed number of ties between CZs o and d , while $S_o \times S_d$ is the product of the number of overall ties in o (S_o) and d (S_d). By 1930, connections extend much more from the Midwest and Great Plains toward California and the Mountain West, while longer links also connect western locations to the Northeast. Connections within the South and from Texas to other regions also become more prominent.

Figure 1: The changing geography and sectoral orientation of family ties



Notes: Shading gives each CZ's manufacturing share among agriculture and manufacturing, with teal denoting lower and orange higher manufacturing shares. Lines show selected strong cross-CZ family connections. Line width increases with relative connectedness, $RC_{od} = S_{od}/(S_o \times S_d)$, and line color gives the manufacturing share among relatives connected across the pair, where S_{od} is the number of ties between CZs o and d , S_o and S_d are their total cross-CZ ties. Striped CZs have no sector-composition data.

Family ties and sectoral choice. I next document a strong positive relationship between workers' sector choices and the sectors in which their local relatives work. Following [Eckert and Peters \(2022\)](#), I concentrate on agriculture and manufacturing to study the transition from the former to the latter sector. However, it is important to keep in mind that services account for a large share of employment: 33.9% in 1880 and 52.6% in 1950.¹⁴ In manufacturing, workers with 5–14 manufacturing relatives in the same CZ are about 7–8 percentage points more likely to work in manufacturing than workers with none. In agriculture, the corresponding difference reaches almost 6 percentage points for workers with 15–19 agricultural relatives.¹⁵ Appendix Figure [A2.3.1](#) plots these relationships after absorbing CZ fixed effects and controlling for individual characteristics and total family size.¹⁶ The gradient is positive in both sectors, and raw employment rates show the same ordering (Appendix Figure [A2.3.2](#)).

These patterns are consistent with family connections affecting sector choice, but they may also reflect persistent differences across families in skills, preferences or other determinants of employment.

Family ties and migration destinations. Family connections are also strongly related to where migrants go. Relatives already established elsewhere can reduce the uncertainty and

¹⁴Industry employment is defined using IPUMS IND1950 codes. Manufacturing includes durable (306–399) and non-durable (406–499) manufacturing, while agriculture includes farming (105).

¹⁵Pairwise family relationships are de-duplicated as described in Section 2, so the same relative is counted only once when reconstructed family trees overlap.

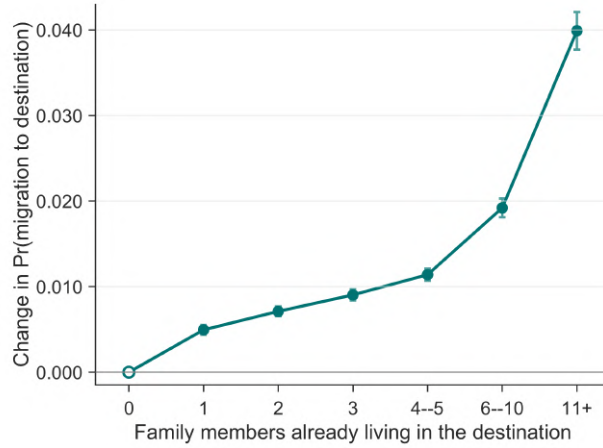
¹⁶Using the share of employed relatives working in the sector yields a similar pattern.

costs of migration through information and temporary support (Carrington, Detragiache, & Vishwanath, 1996; McKenzie & Rapoport, 2007; Munshi, 2003; Stuart & Taylor, 2021), while historical evidence likewise links kinship structure to geographic mobility (Ghosh et al., 2023). I ask whether subsequent migrants disproportionately choose destinations in which members of their extended family already live. For reconstructed family identifier f , origin o , destination d and census t , let n_{fdt} denote the number of family members already living in $d \neq o$ and let y_{fodt} indicate whether at least one member of family f moves from o to d before the next census.¹⁷ I estimate

$$y_{fodt} = \sum_{j \geq 1} \beta_j \mathbf{1}\{n_{fdt} = j\} + \mu_{fot} + \delta_{odt} + \varepsilon_{fodt}, \quad (1)$$

where destinations with no relatives form the omitted category. Family-by-origin-by-census fixed effects μ_{fot} absorb differences in family f 's propensity to send migrants from origin o in census t , while origin-by-destination-by-census fixed effects δ_{odt} absorb common determinants of migration along a given corridor in that period.¹⁸ The specification therefore compares destinations within the same family-origin-period while absorbing common corridor attractiveness.

Figure 2: Migration toward family at destination



Notes: The figure reports estimates from equation (1), with destinations containing no relatives as the omitted category. Standard errors are two-way clustered by reconstructed family identifier and origin–destination corridor. Vertical bars denote 95% confidence intervals.

Figure 2 shows a steep increasing relationship between family presence and subsequent migration. A destination with one relative is already substantially more likely to receive a migrant from the family and the gradient rises with the number of relatives at destination. The corre-

¹⁷Individuals can belong to multiple reconstructed family lineages in this exercise. A given migration can therefore contribute to the lineage-specific outcome of more than one family identifier. Within a family identifier, however, each individual migration is counted only once.

¹⁸To make the destination choice set tractable, I retain all destinations containing relatives and all realized migration destinations, together with twelve randomly sampled zero-kin, non-chosen destinations. Observations are weighted by the inverse of their inclusion probability.

sponding probability for a sampled zero-kin destination is only 0.035%. Although descriptive, the estimates establish a strong directional feature of family connections: conditional on the family’s origin and common corridor attractiveness, subsequent migrants disproportionately move toward destinations where their relatives are already established.

3.2 Deaths as shocks to family connections

The descriptive relationships above do not establish that family connections themselves affect workers’ choices. I therefore use deaths of relatives to generate changes in family presence that are not caused by the focal worker’s own sector or location decision. I study two shocks separately. The first removes a relative employed in a particular sector, while the second removes a local non-employed relative and therefore changes local family presence without simultaneously removing a contemporaneous sector-specific employment connection. Using deaths to generate exogenous changes in social connections follows a longer literature studying collaborators, co-inventors and coworkers (Azoulay, Graff Zivin, & Wang, 2010; Jäger & Heining, 2022; Jaravel, Petkova, & Bell, 2018; Jones & Olken, 2005; Oettl, 2012). Lin and Mo (2024) apply a related design to coworkers’ sectoral connections.

Sector-specific family ties. I exploit the death of a sector-employed relative as a shock to sector-specific family presence. Because deaths are observed only between censuses, I focus on relatives aged 18–40 and employed in the relevant sector in the final census in which they are observed alive. The age restriction reduces the role of predictable old-age mortality and associated anticipatory responses.¹⁹ If the focal individual and deceased relative share multiple family identifiers, their relationship is de-duplicated as in Section 2, so each pair generates at most one treatment relationship.

Identification requires that, conditional on pre-event family characteristics and local labor-market conditions, the timing of an eligible death is unrelated to changes in the individual’s potential sectoral outcomes. I estimate a stacked event study following Wing, Freedman, and Hollingsworth (2024).²⁰ For each treatment cohort, I construct a separate event-time stack around the first eligible death. Event time zero is the first census after the death and the final census before the death, $\tau = -1$, is omitted. Each treatment cohort is compared with individuals who never experience an eligible death and with later-treated individuals whose first eligible death occurs after the end of the cohort’s event window. Individuals who have already been treated, or who become treated within the relevant event window, are excluded

¹⁹Age is measured at the final observation alive rather than at death, which is unavailable. Results are robust to alternative age restrictions.

²⁰Similar dynamic effects can be estimated using the approach of Borusyak, Hull, and Jaravel (2022).

from that stack.^{21,22}

Let $D_{is} = \mathbf{1}\{T_i = e_s\}$, let $r_{ist} \equiv (t - e_s)/10$ denote event time in census decades and let c_{is}^0 denote the individual's CZ in the final pre-event census. For treated individuals, this is also the CZ in which the deceased relative was last observed alive, since eligible deaths require the relative to live in the individual's CZ before death. I estimate

$$\text{ManufEmp}_{ist} = \alpha_{is} + \sum_{\tau \neq -1} \beta_{\tau} D_{is} \mathbf{1}\{r_{ist} = \tau\} + \lambda_{c_{is}^0 ts} + \gamma' A_{it} + \Gamma'_t X_{is}^{\text{pre}} + \varepsilon_{ist}, \quad (2)$$

where α_{is} are individual-by-stack fixed effects and $\lambda_{c_{is}^0 ts}$ are pre-event-CZ-by-year-by-stack fixed effects. Fixing location at its pre-event value absorbs changes in the local labor market without conditioning on subsequent migration, which may itself respond to the family shock. A_{it} contains age controls, while X_{is}^{pre} contains predetermined family and household characteristics interacted with census-year effects. Standard errors are clustered at the de-duplicated family-identifier level.²³

Figure 3 reports the results. Panel (a) shows small pre-event coefficients followed by a persistent decline in manufacturing employment, reaching approximately 1 percentage point three decades after the loss of a manufacturing-employed relative. Relative to a baseline manufacturing share of 36%, this corresponds to a decline of about 2.8%. Panel (b) shows that the response extends beyond immediate household members. Twenty years after the death, manufacturing employment falls by approximately 0.76–0.83 percentage points following the death of a parent, child, sibling or spouse. The response to an uncle or aunt is about 60% of the parent response and the cousin effect is around 0.41 percentage points. The effect therefore declines with genealogical distance but remains visible among extended kin.

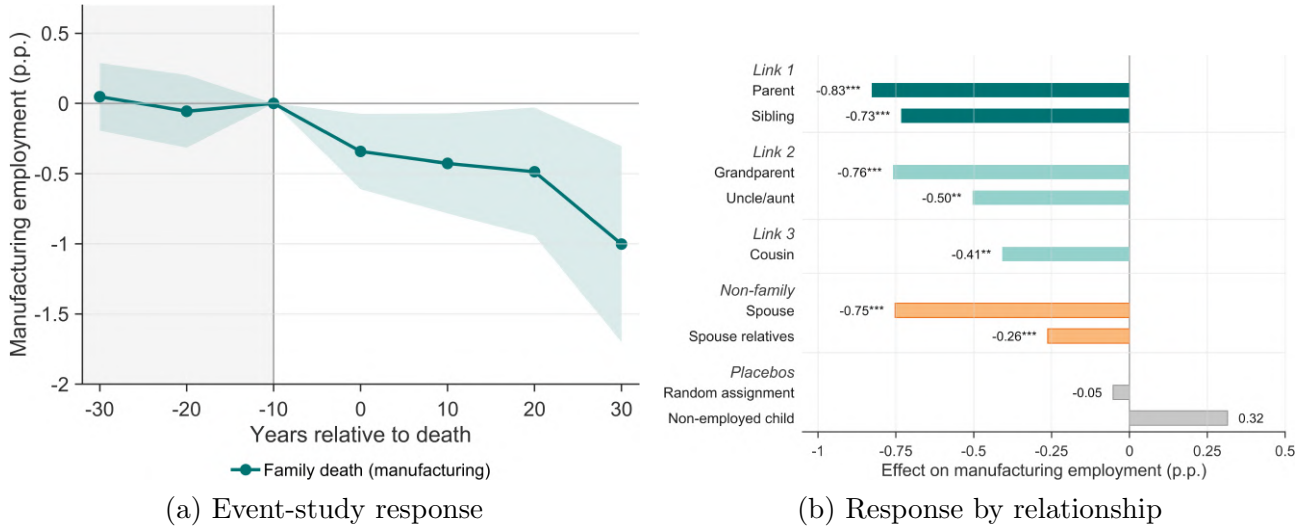
The main threat to interpretation is that the death of a sector-employed relative may affect subsequent sector choice through channels other than the loss of a sector-specific family input. Bereavement, inheritance or transfers of productive assets could also change subsequent employment. Several exercises make these explanations less consistent with the manufacturing results. Deaths of non-employed relatives generate no comparable decline in manufacturing employment, despite exposing families to many of the same consequences of a death. Man-

²¹Each stack includes up to three census waves before and after treatment. Observations are censored from the second eligible death onward. Approximately 11,000 of 1.2 million individuals experience more than one eligible death. The construction ensures that treated observations from other cohorts are never used as controls when treatment effects may vary across cohorts.

²²I apply the corrective stacked-difference-in-differences weights of Wing et al. (2024), which ensure that treated and control observations receive consistent weights across cohort-specific stacks. I multiply these weights by the population weights used to reweight the linked sample toward the full census.

²³If two individuals share multiple family identifiers, the de-duplication rule in Section 2 assigns their relationship to one family identifier before the estimation sample is constructed. Never-treated individuals are assigned the earliest reconstructed family identifier available to them. If multiple identifiers are available at that stage, I retain the identifier associated with the earliest link in its construction and, in the rare remaining ties, the lower family identifier.

Figure 3: Manufacturing employment after the loss of a sector-employed relative



Notes: Panel (a) reports estimates from equation (2); $\tau = -1$ is the final census before the death. Panel (b) reports the manufacturing response twenty years after the first eligible death separately by the focal individual's relationship to the deceased, together with placebo estimates. Shaded regions and vertical bars denote 95% confidence intervals. Standard errors are clustered at the de-duplicated family-identifier level.

Manufacturing deaths are also followed by little change in household composition or fixed-asset holdings. The main exception arises in agriculture, where deaths are associated with lower farm ownership in some specifications.

Mortality could also be related to sectoral conditions. A fatal workplace accident, for example, could both remove a manufacturing-employed relative and convey information about the risks of manufacturing work. Occupational fatalities account for only a small share of mortality over this period. Employment also declines in the deceased relative's sector following deaths in services and agriculture, and randomly reassigning manufacturing deaths across families within the same CZ produces no corresponding effect. This placebo preserves local exposure to manufacturing deaths and local economic conditions while breaking the link between the death and the individual's own family. Appendix A2.3.2 reports these exercises together with robustness to alternative age restrictions, population weights, comparison groups and event-window support.

Local family ties and migration. I next isolate the spatial role of local family connections. I study deaths of non-employed relatives living in the individual's CZ before the death and impose the same age restriction as above. Because the deceased relative is not employed, the shock removes a local family connection without simultaneously removing a contemporaneous employment tie to a particular sector.

I use the same stacked event-study design as above. The outcome is an indicator for changing CZ between consecutive censuses, expressed in percentage points. Because deaths and migration

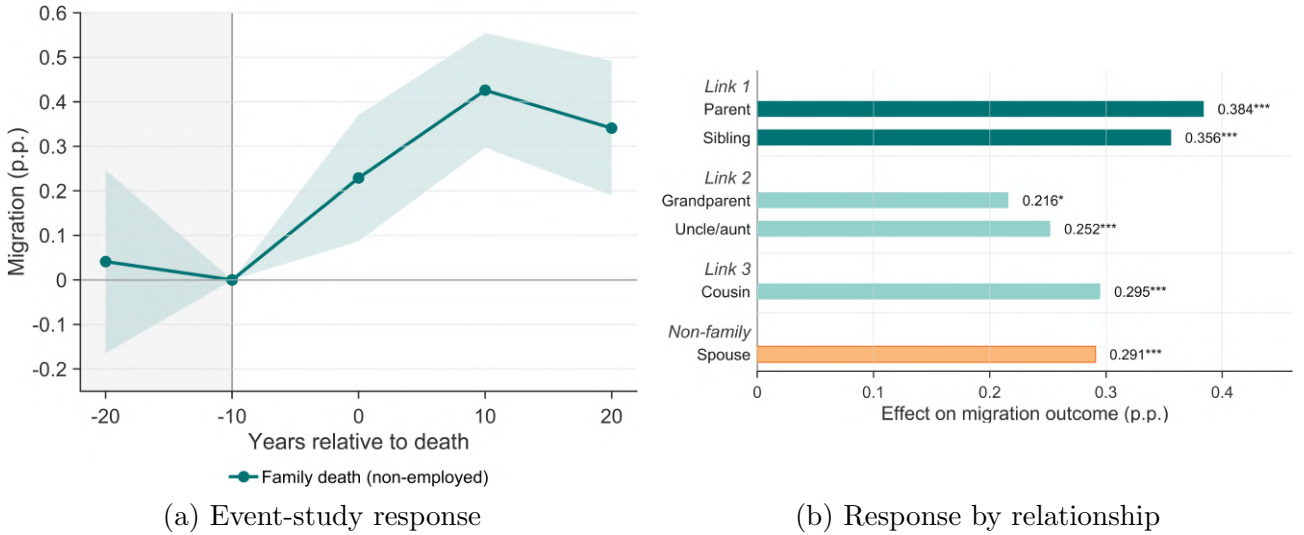
are both observed only between censuses, event time zero corresponds to the migration interval in which the death occurs. I report $\tau = 0$ as part of the event-study path but interpret only $\tau \geq 1$ as clean post-death responses.²⁴

I estimate

$$\text{Mig}_{ist} = \alpha_{is} + \sum_{\tau \neq -1} \beta_{\tau} D_{is} \mathbf{1}\{r_{ist} = \tau\} + \lambda_{c_{is}^0 ts} + \gamma' A_{it} + \Gamma'_t X_{is}^{\text{pre}} + \varepsilon_{ist}, \quad (3)$$

using the same individual-by-stack fixed effects, pre-event-CZ-by-year-by-stack fixed effects, controls and weighting scheme as in equation (2). Standard errors are clustered at the de-duplicated family-identifier level using the same assignment rule as in the sectoral death design.

Figure 4: Migration after the loss of a local family tie



Notes: Panel (a) reports estimates from equation (3); $\tau = -1$ is omitted and $\tau = 0$ is the migration interval in which the death occurs. Panel (b) reports estimated effects on migration separately by relationship, grouped by genealogical distance. The spouse estimate provides a comparison with a non-blood relationship. Shaded regions in Panel (a) denote 95% confidence intervals.

Panel (a) of Figure 4 shows that migration rises after the loss of a local non-employed relative. The effect is approximately 0.43 percentage points in the first clean post-death interval and 0.34 percentage points in the following interval. Relative to a baseline decadal migration rate of 19%, the first response is about 2.3%. Because the deceased was non-employed, the effect does not reflect the direct loss of a contemporaneous sector-specific employment connection. The magnitude is similar to Koenen and Johnston (2024), who find that a friend in the same CZ increases the probability of living there by 0.3 percentage points. Panel (b) shows that the response is not confined to immediate relatives. The effects are positive for parents, siblings,

²⁴The omitted period, $\tau = -1$, is the preceding migration interval and is fully pre-death. Because locations are observed from 1900 onward, the treated migration stacks are the 1920 and 1930 death cohorts. Individuals treated in 1940 can enter earlier stacks as later-treated controls while they remain untreated.

grandparents, aunts and uncles and cousins and remain economically meaningful for more distant kin. The comparable spouse estimate shows that the response also extends beyond blood relationships.

Evidence on mechanisms. The death designs establish that sector-specific and local family connections affect workers' choices, but they do not by themselves identify the mechanisms through which relatives matter. Additional evidence points to several channels. For sector choice, informal contacts can transmit information about jobs and opportunities (Granovetter, 1973; Rees, 1966), while models of social interactions show how such information can generate persistent employment differences (Calvó-Armengol & Jackson, 2004). Using the 1949 *Occupational Outlook Handbook*, Appendix A2.3.3 shows that family occupational persistence is stronger where occupational entry relies primarily on workplace experience than where a formal school-based route exists. Appendix A2.3.4 further shows that employed relatives are associated with lower unemployment and job search, while same-sector family ties predict faster subsequent wage growth among manufacturing workers, especially outside urban labor markets. For migration, the response is consistent with locally provided family support. Proximity to relatives can facilitate childcare and labor supply (Compton & Pollak, 2014) and strengthen attachment to place (Zabek, 2024), while informal insurance can reduce incentives to migrate (Morten, 2019; Munshi & Rosenzweig, 2016). Consistent with these mechanisms, the response to a death of a family member is larger among households with young children (Appendix Table A2.4.1), and local non-employed family presence is associated with greater female employment and less housework (Appendix Figure A2.4.1). I find no corresponding migration response after deaths of relatives outside the individual's CZ and the estimates are similar among individuals without homeownership.

Taken together, the evidence identifies the main margins used in the quantitative model. As structural transformation proceeds, family connections shift toward manufacturing and become more geographically dispersed. Sector-specific and local family presence affect sector choice and migration, while migrants disproportionately choose destinations where relatives already live. The model combines these margins to quantify how workers' choices reshape the connections inherited by later generations.

4 Quantitative model

The previous section highlights the role of kin in reduced form, but its aggregate consequences depend on the joint response of wages, sector choices, migration and kin connections. I therefore embed these mechanisms in a general-equilibrium model that evolves across space and generations.

Building on [Eckert and Kleineberg \(2024\)](#) and [Eckert and Peters \(2022\)](#),²⁵ the model combines non-homothetic demand à la [Boppart \(2014\)](#), trade and local congestion with an endogenous distribution of kin connections across sectors and space. Kin affect choices through two channels. Sector-specific connections lower the utility cost of entering a sector, capturing information, referrals and know-how documented in Section 3, while nearby kin raise location utility through forms of home support such as childcare ([Compton & Pollak, 2014](#)) and place attachment as in [Zabek \(2024\)](#). These choices then reshape the sectoral and spatial distribution of kin inherited by later cohorts. Lastly, parents also value their children’s future opportunities, introducing an additional intergenerational link ([Barro & Becker, 1989](#); [Eckert & Kleineberg, 2024](#); [Fogli & Guerrieri, 2019](#)).

Environment. The economy has a finite set of regions $n \in N$, with $|N| = 455$ corresponding to commuting zones and two sectors $k \in \{A, M\}$. Time is discrete and one period is a 20-year interval, so that each model period corresponds approximately to one biological generation, as in [Eckert and Peters \(2022\)](#) and [Eckert and Kleineberg \(2024\)](#).²⁶ The population is organized in overlapping generations. Adults work, consume and choose where to live, while their children choose a sector before adulthood that remains fixed over the adult life cycle. This approximation separates sectoral from spatial adjustment within a generation. It is motivated by [Eckert and Peters \(2022\)](#), who show that much historical sectoral reallocation occurred within locations, and by Appendix A2.4.3, where migration induced by family deaths produces no detectable sectoral switching.

Generations are linked through a stylized kin structure. Each adult belongs to a parental branch that gives rise to a mass ϕ of children in the next generation. These children form a sibling group. When they become adults, their sector and location choices determine where the members of that sibling group live and work. From the perspective of their children, the other members of the group are aunts and uncles. I refer to these non-parental relatives as horizontal kin. Each worker therefore inherits two types of family connections from the previous generation: one parental connection and a distribution of horizontal kin across sectors and locations.

Population grows across generations at the common gross rate ϕ , which I impose equally across ancestral-origin groups.²⁷ Tracking the complete family tree behind this structure would nevertheless require carrying forward the histories of many relatives and potentially their simulta-

²⁵These papers are part of a broader quantitative spatial literature studying the dynamics of development, trade and labor reallocation ([Caliendo, Dvorkin, & Parro, 2019](#); [Desmet, Nagy, & Rossi-Hansberg, 2018](#)); see [Redding and Rossi-Hansberg \(2017\)](#) for a review.

²⁶Model periods are spaced 20 years apart except between 1880 and 1910, where the destruction of the 1890 census creates a 30-year interval.

²⁷The common growth rate ensures that differences in kin mass across ancestral-origin groups do not mechanically generate differences in population growth or fertility.

neous choices. I impose four restrictions that reduce this dimensionality while preserving the intergenerational kin mechanism.

Assumption 1. *Families are grouped by a permanent ancestral-origin type g , defined by the commuting zone of their linked 1850 ancestor.*

Rather than following each family separately, the model tracks the average distribution of kin across sectors and locations among families with the same g . This preserves persistent differences in kin geography across ancestral origins while abstracting from differences across individual families within the same origin group. Appendix A4.7 shows that the approximation preserves the aggregate transition reasonably well despite heterogeneity within g .

Assumption 2. *Workers' sector and location choices depend only on kin from the previous generation: their parent and the parent's siblings.*

This makes the kin environment predetermined when a worker chooses sector and location and limits the family history that must be carried forward to one generation. The cost is that the model does not capture direct effects from siblings, cousins or other relatives choosing in the same generation, but it substantially reduces history dependence and the state space of the model and avoids contemporaneous family complementarities that could generate additional multiplicity.

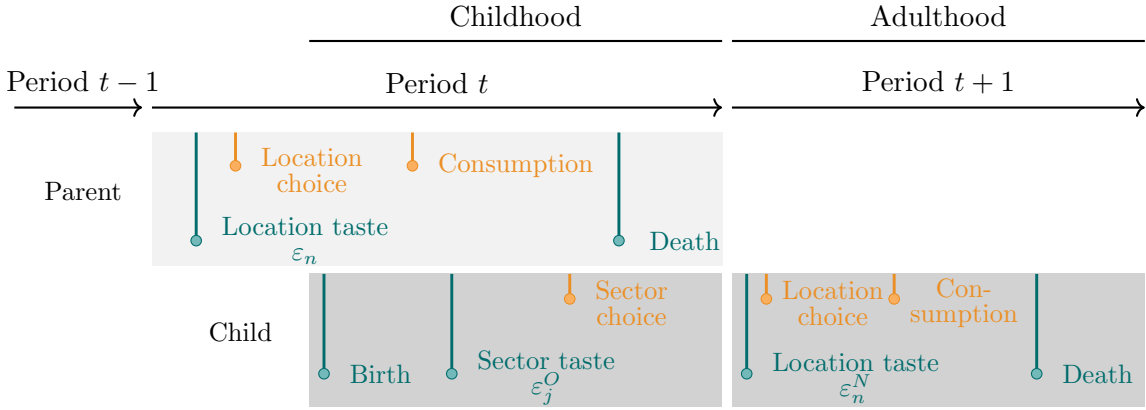
Assumption 3. *For each ancestral-origin type g , total horizontal-kin mass r_g is fixed over time.*

This assumption isolates changes in the composition of kin from changes in the number of relatives. The reconstructed data do not cleanly identify changes in kin quantity over time (Table 1), although U.S. fertility declined substantially over this period: the total fertility rate fell from about 4.2 births per woman in 1880 to about 3.0 in 1950. I therefore allow the sectoral and spatial distribution of kin to evolve while holding total kin mass r_g fixed. Population itself is not fixed and continues to grow at rate ϕ . Allowing kin mass to vary could strengthen or weaken persistence depending on where those relatives live and work. Appendix A5.4 shows that the main qualitative results are robust to alternative levels of horizontal-kin mass.

Assumption 4. *Conditional on their inherited kin environment, workers make independent choices under Type-I extreme-value sector and location shocks.*

The shocks are independent across workers, alternatives and periods, with mean zero and dispersion parameters σ_O for sectors and σ_N for locations. Independence makes individual choice probabilities sufficient to determine how a sibling group is distributed across sectors and locations, rather than requiring its joint choices, while the extreme-value assumption delivers closed-form choice probabilities and expected values. The cost is that the model abstracts from direct coordination and correlated idiosyncratic choices among relatives.

Figure 5: Model timeline



Notes: A parent enters period t , chooses a location and consumes. Children are born in t , draw sector taste shocks and choose sectors before adulthood. As adults in $t+1$, they draw location shocks and choose where to live, becoming the parents of the next generation. Teal markers denote shocks and orange markers denote endogenous choices.

Together, these restrictions allow the model to carry forward only the parental connection and the expected distribution of horizontal kin for each ancestral-origin type. As one generation becomes the next, workers' sector and location choices reshape that distribution. Entry into manufacturing increases the mass of manufacturing kin inherited by the next generation, while migration increases kin presence at the destination. The model can therefore capture how current choices change the family connections faced by later generations without tracking complete family trees. Figure 5 summarizes the timing.

Kin distribution. I summarize horizontal kin by their overall mass and their distribution across sectors and locations. Consider an adult of ancestral type g who lives in location m and works in sector j . Let $\Phi_{g,t}(m, j; n, k)$ denote the mass of that adult's horizontal kin located in region n and employed in sector k .

I write this mass as

$$\Phi_{g,t}(m, j; n, k) = r_g \lambda_{g,t}(n, k \mid m, j),$$

where $\lambda_{g,t}(n, k \mid m, j)$ is the share of horizontal kin in location n and sector k , conditional on the adult's current state (m, j) , and these shares sum to one across all locations and sectors. The scalar r_g is the total horizontal-kin mass associated with one adult of ancestral type g , disciplined by the average number of reconstructed sibling links per adult in that ancestral-origin group.²⁸ The term $\lambda_{g,t}(n, k \mid m, j)$ is the share of that total mass located in (n, k) , conditional on the adult being in (m, j) .

²⁸For example, a focal individual has around 1.4 aunts/uncles in 1880, so $r_g = 1.4$.

Given Assumptions 3 and 4, total horizontal-kin mass remains fixed while its distribution can evolve. Conditional on their inherited kin environment, siblings face the same systematic choice problem but draw independent idiosyncratic shocks. Their joint choice distribution therefore factorizes into individual choice probabilities, so the model updates the expected kin shares $\lambda_{g,t}$ using the sector and location choice probabilities derived below.

Consumption and structural transformation. Consumption follows PIGL preferences of Boppart (2014), as used by Eckert and Peters (2022). For nominal wages w and prices $p_{nt} = (p_{nAt}, p_{nMt})$, indirect utility and the agricultural expenditure share are

$$v(w, p_{nt}) = \frac{1}{\epsilon} \left(\frac{w}{p_{nAt}^\varphi p_{nMt}^{1-\varphi}} \right)^\epsilon - \nu \log \left(\frac{p_{nAt}}{p_{nMt}} \right), \quad (4)$$

$$\delta_A(w, p_{nt}) = \varphi + \nu \left(\frac{w}{p_{nAt}^\varphi p_{nMt}^{1-\varphi}} \right)^{-\epsilon}. \quad (5)$$

with

$$\delta_M(w, p_{nt}) = 1 - \delta_A(w, p_{nt}). \quad (6)$$

Because $\epsilon > 0$, the agricultural expenditure share falls with income, generating the non-homothetic Engel effects that drive structural change: as a region grows richer, demand tilts toward the manufacturing good, raising its relative wage and the return to entering it.

Children's sector choice. Before adulthood, a child of ancestral type g whose parent is in (n, j) chooses a sector, trading off a non-kin entry cost, sector-specific kin connections and the lifetime value associated with that sector. The systematic value of choosing sector k is

$$\mathcal{U}_{gnjt}^k = Q_{gnjt}^k - \tau_{gnjt}^k + V_{t+1}(n, k, g), \quad (7)$$

where τ_{gnjt}^k is a non-kin entry cost, Q_{gnjt}^k is the kin contribution and $V_{t+1}(n, k, g)$ is the continuation value from the adult location problem defined below.

The term Q_{gnjt}^k captures the sector-specific connections available to the child. Entering sector k may require information about jobs and employers as well as sector-specific skills or practical knowledge. Relatives already working in the sector can reduce these barriers through referrals, employer contacts and workplace experience (Calvó-Armengol & Jackson, 2004; Ioannides & Loury, 2004; Montgomery, 1991). I therefore model kinship as reducing non-wage barriers to sector entry, as in Takeda (2022).

Let $\Phi_{gnjt}^k(m) \equiv \Phi_{g,t}(n, j; m, k)$ denote the horizontal kin mass living in location m and working

in sector k . I combine parental and horizontal kin into an effective stock of sector- k connections:

$$F_{gnjt}^k = \mathbf{1}\{j = k\} + \psi_R^k \Phi_{gnjt}^k(n), \quad (8)$$

where the parent's contribution is normalized to one and ψ_R^k measures the contribution of one unit of local horizontal kin mass in sector k relative to the parent. The kin contribution to sector choice is $Q_{gnjt}^k = \theta_k F_{gnjt}^k$, where θ_k governs how strongly sector choice responds to these connections.

The child chooses the sector that maximizes $\mathcal{U}_{gnjt}^k + \sigma_O \varepsilon_k^O$, where ε_k^O is the sector taste shock from Assumption 4. Assumption 4 then implies the sector-choice probabilities and expected value

$$P_t^O(k | n, j, g) = \frac{\exp(\mathcal{U}_{gnjt}^k / \sigma_O)}{\sum_{\ell \in \{A, M\}} \exp(\mathcal{U}_{gnjt}^\ell / \sigma_O)}, \quad (9)$$

$$O_t(n, j, g) = \sigma_O \log \sum_{\ell \in \{A, M\}} \exp(\mathcal{U}_{gnjt}^\ell / \sigma_O). \quad (10)$$

The opportunity value $O_t(n, j, g)$ is the child's expected value before the sector taste shocks are realized and enters the parent's location choice below.

Parents' location choice. Adults choose where to live as a dynamic discrete choice, internalizing both their own utility and, through altruism, their children's future opportunities. This forward-looking migration problem follows the dynamic discrete-choice model of [Kennan and Walker \(2011\)](#) and is embedded here in a quantitative spatial environment related to [Fogli and Guerrieri \(2019\)](#) and [Eckert and Kleineberg \(2024\)](#).

Consider an adult of ancestral type g , whose origin is n and whose own fixed sector is j , who chooses destination m . The worker earns wage w_{mt}^j . Let $\mathcal{A}_{gmt}^j$ denote the pre-existing non-kin amenity of destination m for this worker. The systematic value of destination m is

$$B_{gmt}^j(m) = \mathcal{A}_{gmt}^j - \rho \log L_{mt} + v(w_{mt}^j, p_{mt}) + \theta_H H_{gnjt}(m) - M_{gnmt}^j + \beta O_t(m, j, g), \quad (11)$$

where L_{mt} is total employment at the destination, M_{gnmt}^j is the migration cost, $H_{gnjt}(m)$ is the kin amenity defined below and $\beta < 1$ is parental altruism. The final term, $O_t(m, j, g)$, is the value of the sectoral opportunities that destination m confers on the worker's children.

Kin proximity can facilitate childcare and female labor supply ([Compton & Pollak, 2014](#)), while related evidence shows that family peers affect mothers' labor-supply choices ([Nicoletti, Salvanes, & Tominey, 2018](#)). I capture these forms of family support as a location amenity.

Under Assumption 2, an adult's location choice depends on kin inherited from the previous

generation: the parent and the parent's siblings. The parent provides a connection to the location where the adult grew up, while the parent's siblings—the adult's aunts and uncles—provide connections to the places where they live. The relevant kin geography is therefore predetermined when the adult chooses where to live.

Consider an adult of ancestral type g , who grew up in location n and works in sector j in period t . The locations of the adult's aunts and uncles are given by the parent's sibling distribution in period $t - 1$. This distribution depends on the parent's sector q , which need not coincide with the adult's own sector j . Because the parent's sector is not part of the adult's current state (n, j, g) anymore, I average over possible parental sectors using their posterior probabilities. The probability that the parent worked in sector q , conditional on the adult's current state, is

$$\Pr_{t-1}(q \mid n, j, g) = \frac{L_{gn,t-1}^q P_{t-1}^O(j \mid n, q, g)}{\sum_{\tilde{q} \in \{A, M\}} L_{gn,t-1}^{\tilde{q}} P_{t-1}^O(j \mid n, \tilde{q}, g)}.$$

The kin amenity associated with destination m is then

$$H_{gnjt}(m) = \mathbf{1}\{m = n\} + \psi_R^H \sum_{q \in \{A, M\}} \Pr_{t-1}(q \mid n, j, g) \sum_{\ell \in \{A, M\}} \Phi_{gnq,t-1}^\ell(m), \quad (12)$$

where the first term captures the parental connection to the childhood location n . The second captures the expected mass of aunts and uncles living in destination m , averaging over the parent's possible sector. Because location utility depends on the total kin mass at a destination rather than the sectors in which relatives work, I sum this mass across sectors ℓ . The parameter ψ_R^H measures the contribution of one unit of horizontal kin mass relative to the parent.

The worker chooses the destination that maximizes $B_{gnt}^j(m) + \sigma_N \varepsilon_m^N$, where ε_m^N is the location taste shock from Assumption 4. Assumption 4 then implies the location-choice probabilities and expected value

$$P_t^N(m \mid n, j, g) = \frac{\exp(B_{gnt}^j(m)/\sigma_N)}{\sum_{r \in N} \exp(B_{gnt}^j(r)/\sigma_N)}, \quad (13)$$

$$V_t(n, j, g) = \sigma_N \log \sum_{m \in N} \exp(B_{gnt}^j(m)/\sigma_N). \quad (14)$$

The parameter σ_N governs how sharply households sort on differences in location value. A higher σ_N makes location choices less sensitive to changes in wages, amenities and migration costs. Equations (13)–(14) use j for an adult's own sector. Thus, for a child who selected sector k in equation (9), the relevant location probability is $P_{t+1}^N(m \mid n, k, g)$: the parent's sector affects the child's sector choice, after which the child's own sector becomes the state for

migration.

Evolution of kin connections. Sector and location choices determine how horizontal kin are distributed for the next generation. I now use the choice probabilities defined above to construct the law of motion for $\lambda_{g,t}$.

Consider descendants of adults of ancestral type g in state (m, j) in period t . A fraction $P_t^O(k | m, j, g)$ enters sector k . Once sector k is chosen, it becomes the descendant's own sectoral state and a fraction $P_{t+1}^N(n | m, k, g)$ subsequently chooses location n . Hence, descendants from parental state (m, j) reach state (n, k) with their joint probability.

Adults observed in the same state (n, k) at $t + 1$ can descend from different parental states. The share of type- g adults in (n, k) whose parent was in (m, j) is therefore

$$\Pr_t(m, j | n, k, g) = \frac{L_{gmt}^j P_t^O(k | m, j, g) P_{t+1}^N(n | m, k, g)}{\sum_{\tilde{m} \in N} \sum_{\tilde{j} \in \{A, M\}} L_{g\tilde{m}t}^{\tilde{j}} P_t^O(k | \tilde{m}, \tilde{j}, g) P_{t+1}^N(n | \tilde{m}, k, g)}. \quad (15)$$

This is a posterior over parental states within the current (g, n, k) state.²⁹

Conditional on parental state (m, j) , the horizontal-kin endpoint is assigned the same sector and location choice distribution as descendant mass generated by that state. It reaches (n', k') with probability $P_t^O(k' | m, j, g) P_{t+1}^N(n' | m, k', g)$. Averaging this distribution over the parental histories represented among adults currently in (n, k) gives

$$\lambda_{g,t+1}(n', k' | n, k) = \sum_{m \in N} \sum_{j \in \{A, M\}} \Pr_t(m, j | n, k, g) P_t^O(k' | m, j, g) P_{t+1}^N(n' | m, k', g). \quad (16)$$

The corresponding horizontal kin mass in the next generation is

$$\Phi_{g,t+1}(n, k; n', k') = r_g \lambda_{g,t+1}(n', k' | n, k). \quad (17)$$

Since both the posterior weights and the choice probabilities sum to one, the law of motion reallocates the fixed total horizontal-kin mass r_g across sectors and locations, as imposed in Assumption 3.

Equation (16) is a conditional-mean law of motion. The model pools distinct genealogical lineages that share the same ancestral type and current state and propagates their average kin distribution. This is the dimensionality reduction that makes the endogenous evolution of kin tractable without carrying individual family trees through the state space.

²⁹The common cohort-growth factor ϕ scales descendants from every parental state equally and therefore cancels from equation (15).

Production and trade. Sectoral wages and prices are determined in general equilibrium through a standard Eaton–Kortum trade structure (Eaton & Kortum, 2002). Embedding this trade block in the dynamic spatial economy follows the broader approach of Caliendo et al. (2019). Each good is produced by competitive firms using labor,

$$q_{nt}^k = z_{nt}^k L_{nt}^k,$$

with Fréchet productivity draws of scale Z_{nt}^k and shape μ_k . Trade is subject to iceberg costs T_{nmt}^k , where n denotes the importing location and m the producing location. Because labor is the only input, unit production cost is the local wage. The expenditure share and sectoral price index are therefore

$$\pi_{nmt}^k = \frac{(w_{mt}^k T_{nmt}^k)^{-\mu_k} (Z_{mt}^k)^{\mu_k}}{\sum_{r \in N} (w_{rt}^k T_{nrt}^k)^{-\mu_k} (Z_{rt}^k)^{\mu_k}}, \quad (18)$$

$$p_{nt}^k = \Gamma^k \left[\sum_{m \in N} (w_{mt}^k T_{nmt}^k)^{-\mu_k} (Z_{mt}^k)^{\mu_k} \right]^{-1/\mu_k}. \quad (19)$$

Aggregation and market clearing. The labor distribution evolves in parallel with the kin distribution. Let L_{gnt}^j denote adults of ancestral type g in sector j and location n at time t . Applying the sector and location choices to the entire adult population and scaling the next cohort by the common cohort-growth factor ϕ gives

$$L_{gm,t+1}^k = \phi \sum_{n \in N} P_{t+1}^N(m | n, k, g) \sum_{j \in \{A, M\}} P_t^O(k | n, j, g) L_{gnt}^j. \quad (20)$$

Thus, ϕ scales aggregate adult cohort growth, while the choice probabilities determine how each ancestral group is reallocated across sectors and locations.

Aggregate sectoral and local employment are

$$L_{nt}^k = \sum_g L_{gnt}^k, \quad L_{nt} = \sum_{k \in \{A, M\}} L_{nt}^k.$$

Labor markets clear location by location and goods markets clear using the PIGL expenditure shares:

$$w_{nt}^k L_{nt}^k = \sum_{m \in N} \pi_{mnt}^k X_{mt}^k, \quad (21)$$

$$X_{nt}^k = \sum_g \sum_{j \in \{A, M\}} \delta_k(w_{nt}^j, p_{nt}) w_{nt}^j L_{gnt}^j. \quad (22)$$

The first condition equates the value of production in sector k and location n to expenditures on goods produced there across all importing locations. The second aggregates sector- k expenditure across households residing in location n .

Equilibrium. The endogenous state of the economy at date t is given by the distribution of workers across ancestral types, sectors and locations, L_t , together with the kin distribution, Φ_t . Let $S_t \equiv (L_t, \Phi_t)$ collect these endogenous state variables and let Ω_t denote the exogenous fundamentals of the economy, including productivity, trade costs, non-kin amenities, sector-entry costs and migration costs. Following [Caliendo et al. \(2019\)](#) and [Eckert and Kleineberg \(2024\)](#), I define the equilibrium in two steps.

Temporary equilibrium. Conditional on the state S_t and current fundamentals Ω_t , a temporary equilibrium is a vector of wages, prices, trade shares and expenditures $\{w_t, p_t, \pi_t, X_t\}$ that satisfies the production and trade conditions in equations (18)–(19) and the market-clearing conditions in equations (21)–(22). Thus, conditional on the current labor distribution, the temporary equilibrium determines the within-period allocation and factor prices. The kin distribution Φ_t does not enter the static market-clearing conditions directly, but it is part of the dynamic choices because it affects sector and location choices.

Sequential competitive equilibrium. Given an initial state $S_0 = (L_0, \Phi_0)$, a path of fundamentals $\{\Omega_t\}_{t=0}^\infty$ and the structural parameters, a sequential competitive equilibrium is a sequence

$$\{V_t, O_t, P_t^N, P_t^O, L_t, \Phi_t, w_t, p_t, \pi_t, X_t\}_{t=0}^\infty$$

such that, at every date t : (i) sector choices satisfy equations (7)–(10); (ii) location choices satisfy equations (11)–(14); (iii) wages, prices, trade shares and expenditures constitute a temporary equilibrium given S_t and Ω_t ; (iv) the labor distribution evolves according to equation (20); and (v) the kin distribution evolves according to equations (16)–(17).

Implications. To make the model’s dynamic feedback transparent, consider a two-sector, two-location version for one ancestral group g , with origin n and destination m . Write $P_t^{O,M} \equiv P_t^O(M \mid n, j, g)$ for the probability of entering manufacturing and $P_{t+1}^{N,k} \equiv P_{t+1}^N(m \mid n, k, g)$ for the probability of moving to m conditional on sector k . Suppressing the parental state, let $\Phi_t^{\ell k}$ denote horizontal-kin mass in location $\ell \in \{n, m\}$ and sector $k \in \{A, M\}$, with total mass r_g , and let $\Phi_t^\ell = \Phi_t^{\ell A} + \Phi_t^{\ell M}$. Appendix [A3.1](#) provides more detail.

Inherited kin shape current choices. The sector and location choices imply

$$\log \frac{P_t^O(M \mid n, j, g)}{P_t^O(A \mid n, j, g)} = a_t^O + \frac{\theta_M \psi_R^M}{\sigma_O} \Phi_t^{nM} - \frac{\theta_A \psi_R^A}{\sigma_O} \Phi_t^{nA},$$

and

$$\log \frac{P_{t+1}^N(m \mid n, k, g)}{P_{t+1}^N(n \mid n, k, g)} = a_{t+1}^{N,k} + \frac{\theta_H \psi_R^H}{\sigma_N} (\Phi_t^m - \Phi_t^n),$$

where the a -terms collect the remaining sector and location values. Manufacturing and destination kin therefore favor the corresponding transitions, while agricultural and origin kin favor persistence. One additional local manufacturing connection raises sector-choice log odds by $\theta_M \psi_R^M / \sigma_O$, while one additional local agricultural connection lowers them by $\theta_A \psi_R^A / \sigma_O$. Similarly, destination kin raise migration log odds while origin kin lower them.

Inheritance changes responses to new opportunities. Consider a temporary increase in the manufacturing wage at the origin, $\log w_{n,t+1}^M = \log \bar{w}_n^M + \varepsilon_t$, with $\varepsilon_{t+1} = 0$, and define $\chi_{t+1}^M \equiv \partial v(w_{n,t+1}^M, p_{n,t+1}) / \partial \log w_{n,t+1}^M > 0$. The shock raises manufacturing entry but, conditional on manufacturing, makes remaining at the origin more attractive. The two responses are

$$R_t^O \equiv \frac{\partial P_t^{O,M}}{\partial \varepsilon_t} = \frac{\chi_{t+1}^M}{\sigma_O} \left(1 - P_{t+1}^{N,M}\right) P_t^{O,M} \left(1 - P_t^{O,M}\right) > 0,$$

and

$$R_t^N \equiv -\frac{\partial P_{t+1}^{N,M}}{\partial \varepsilon_t} = \frac{\chi_{t+1}^M}{\sigma_N} P_{t+1}^{N,M} \left(1 - P_{t+1}^{N,M}\right) > 0.$$

Inherited kin therefore affect not only choice levels but also responsiveness to new opportunities. Because a binary logit responds in proportion to $P(1-P)$, connections that reinforce an initially common choice move its probability farther from one half and reduce its responsiveness to changes in fundamentals. Early in structural transformation, when agriculture and staying at the origin are common, agricultural kin therefore dampen the manufacturing response, while origin kin dampen the migration response in absolute value. Manufacturing and destination kin can have the opposite effect by raising the initially low probabilities of entering manufacturing and moving. This static attenuation mechanism is closely related to that in [Koenen and Johnston \(2024\)](#).

Current choices propagate kin connections. The same choices determine the kin environment inherited by the next generation. In particular, local manufacturing kin satisfy

$$\Phi_{t+1}^{nM} = r_g P_t^{O,M} \left(1 - P_{t+1}^{N,M}\right),$$

while total destination kin are

$$\Phi_{t+1}^m = r_g \left[P_t^{O,M} P_{t+1}^{N,M} + (1 - P_t^{O,M}) P_{t+1}^{N,A} \right].$$

The wage shock therefore creates more local manufacturing kin:

$$\frac{\partial \Phi_{t+1}^{nM}}{\partial \varepsilon_t} = r_g \left[\left(1 - P_{t+1}^{N,M}\right) R_t^O + P_t^{O,M} R_t^N \right] > 0,$$

and, correspondingly, fewer local agricultural kin. Its effect on destination kin is instead ambiguous:

$$\frac{\partial \Phi_{t+1}^m}{\partial \varepsilon_t} = r_g \left[\left(P_{t+1}^{N,M} - P_{t+1}^{N,A}\right) R_t^O - P_t^{O,M} R_t^N \right].$$

The first force reflects additional manufacturing entry and the second the fact that the higher origin wage discourages manufacturing workers from migrating. Sectoral and spatial propagation can therefore reinforce or offset one another.

Inheritance and propagation determine the cumulative effect of kin. Neither force has a fixed sign. When most workers initially remain in agriculture and at their origin, additional kin in agriculture or at the origin make these already common choices even more likely. This can dampen adjustment toward manufacturing and new destinations. Kin working in manufacturing or living at the destination can instead raise the initially low probability of making these transitions and increase responsiveness. As workers switch sectors and migrate, they also create connections to new sectors and places for other kin and later generations. Propagation therefore amplifies adjustment when these new connections point toward the sectors and places to which later generations reallocate and attenuates it when they point in the opposite direction.

In the quantitative analysis, I therefore distinguish *inheritance*, the effect of the initial kin distribution on reallocation, from *propagation*, the additional effect generated as workers' choices reshape the connections inherited by other kin and later generations.

Propagation creates spillovers across kin. These effects are not fully internalized by the worker making the initial choice. Through continuation values and parental altruism, workers take into account the consequences of their choices for their own descendants. They do not take into account how the same choices change the connections inherited by their siblings' children. A worker who enters manufacturing becomes a manufacturing connection for nieces and nephews, while a worker who migrates becomes a destination connection for them.

This is the kin externality in the model. Its strength is governed by the same parameters estimated above: $\theta_k \psi_R^k / \sigma_O$ determines the effect of a horizontal sector connection, $\theta_H \psi_R^H / \sigma_N$ the effect of a horizontal location connection and r_g the amount of horizontal kin exposure through which these effects propagate. Because total horizontal kin mass is fixed, moving one horizontal connection from the origin to the destination both removes an origin connection and creates a destination connection, changing destination-versus-origin log odds by $2\theta_H \psi_R^H / \sigma_N$.

Extensions. The baseline isolates kin effects through sector entry and location amenities. Three additional patterns motivate extensions: workers with same-sector kin experience faster wage growth, nearby relatives can provide home-production support and schooling partly substitutes for kin in sector entry. The appendix captures these patterns through kin wage learning (Appendix A3.2.1), home support that also raises effective labor supply (Appendix A3.2.2) and schooling as a formal substitute for kin-based sector entry. These additional channels are quantitatively secondary for the main results and hence excluded in the baseline results.

5 Model quantification

I quantify the model for the U.S. economy at four dates between 1880 and 1950: 1880, 1910, 1930 and 1950, with 1970 as the terminal period. I proceed sequentially. First, I take demand, trade, congestion and altruism parameters from external sources and initialize labor and kin distributions from the data. Second, deaths of parents and aunts or uncles identify the strength of kin connections relative to the dispersion in idiosyncratic sector and location preferences. Third, a migration-gravity equation identifies migration costs, while the static trade block recovers productivity. Fourth, I estimate the dispersion parameters governing sector and location choice by indirect inference using quasi-experimental variation from the Dust Bowl, an agricultural productivity shock. Finally, I invert the model to recover the remaining non-kin amenities and sector-entry costs. Table 2 summarizes the resulting parameterization.

5.1 External parameters and initialization

External parameters. I take the externally disciplined parameters from the literature in three steps. First, the parameters governing non-homothetic demand are taken from Eckert and Peters (2022), while parental altruism is disciplined using Eckert and Kleineberg (2024). Second, trade costs are constructed from the historical county-to-county transportation-cost data of Hornbeck and Rotemberg (2024) and the trade elasticities for agriculture and manufacturing are taken from Yang (2024). Appendix A4 describes the aggregation of transportation costs to commuting zones. Third, population congestion is set to the value used by Eckert and Peters (2022), following Allen and Donaldson (2020).

Initial kin distribution. I initialize the labor distribution from the Census and the horizontal kin distribution from the linked genealogical data. The latter requires two objects: the total amount of horizontal kin associated with each ancestral type and the distribution of those relatives across sectors and locations.

For an adult i of ancestral-origin type g , let $R_{i0}(n, k)$ denote the number of reconstructed and

de-duplicated siblings of i who live in region n and work in sector k in 1860.³⁰ These siblings become aunts and uncles from the perspective of the adult's children. I measure the average horizontal-kin mass for ancestral type g as

$$r_g = \mathbb{E} \left[\sum_{n \in N} \sum_{k \in \{A, M\}} R_{i0}(n, k) \middle| g(i) = g \right]. \quad (23)$$

This equation averages the total number of reconstructed horizontal relatives across all individuals with ancestral origin g , regardless of where those relatives live or in which sector they work.

I then use the data to determine where these kin live and in which sectors they work. For adults of type g observed in state (m, j) , define

$$\lambda_{g,0}(n, k \mid m, j) = \frac{\mathbb{E}[R_{i0}(n, k) \mid g(i) = g, (m_i, j_i) = (m, j)]}{\sum_{n' \in N} \sum_{k' \in \{A, M\}} \mathbb{E}[R_{i0}(n', k') \mid g(i) = g, (m_i, j_i) = (m, j)]}. \quad (24)$$

This equation gives the share of horizontal kin located in each region-sector cell (n, k) among adults with ancestral origin g who themselves live in m and work in j . Thus, r_g determines the total horizontal-kin mass, while $\lambda_{g,0}(n, k \mid m, j)$ determines how that mass is distributed across sectors and locations. By construction, these shares sum to one.

Consistent with Assumption 3 in Section 4, the model assigns the same total horizontal-kin mass r_g to all current states within ancestral type g . The initial model-implied kin stock is therefore

$$\Phi_{g,0}(m, j; n, k) = r_g \lambda_{g,0}(n, k \mid m, j).$$

I hold r_g fixed thereafter and use equation (16) to evolve the sectoral and geographic composition of horizontal kin endogenously.

Cohort growth. I use the initial average horizontal degree to discipline common cohort growth, separately from the evolution of kin composition. Let $\bar{r}$ denote the initial-population-weighted average of r_g across ancestral types. I set the common gross cohort-growth factor to

$$\phi = \frac{1 + \bar{r}}{2}. \quad (25)$$

The term $1 + \bar{r}$ is the implied average surviving sibling-group size, including the individual themselves. Dividing by two converts the sibling-group size generated by a parental pair into gross cohort growth per adult. In the data, $\bar{r} = 1.4$, implying $\phi = 1.2$. This fits the overall population growth rate well during this time period. The same ϕ applies to all ancestral

³⁰The same is done for the uncles/aunts for parents in 1880.

types, so differences in r_g do not themselves generate differential population growth. Finally, I initialize $P_{1860}^O(\cdot)$ using observed sector choices between 1860 and 1880.

Table 2: Parameterization of the quantitative model

Symbol	Description	Value	Discipline
<i>A. External and calibrated parameters</i>			
μ_A, μ_M	Trade elasticities	$\{12, 6.5\}$	Yang (2024)
ϵ	Engel elasticity	0.8	Eckert and Peters (2022)
ν	PIGL parameter	1	Eckert and Peters (2022)
φ	Asymptotic agricultural share	0.01	Eckert and Peters (2022)
ρ	Congestion	0.15	Allen and Donaldson (2020)
β	Parental altruism	0.231	Eckert and Kleineberg (2024)
ϕ	Cohort growth	1.2	Eq. (25)
<i>B. Idiosyncratic-choice dispersion</i>			
σ_N	Location taste dispersion	0.51	Indirect inference
σ_O	Sector taste dispersion	0.60	Indirect inference
<i>C. Kin parameters</i>			
ψ_R^H	Aunt/uncle location weight	0.874	Death ratio, Eq. (29)
$\Theta_H \equiv \theta_H / \sigma_N$	Parent location value / dispersion	0.024	Parent-death effect, Eq. (29)
ψ_R^A, ψ_R^M	Aunt/uncle sector weights	$\{0.305, 0.141\}$	Death ratios, Eq. (31)
Λ_A, Λ_M	Parent sector values / dispersion	$\{0.0206, 0.0489\}$	Parent-death effects, Eq. (31)
<i>D. Recovered fundamentals</i>			
$\mathcal{A}_{gmt}^j$	Location-sector amenities	—	Model inversion
Z_{nt}^k	Productivity	—	Trade inversion
τ_{gnjt}^k	Sector-entry costs	—	Model inversion
M_{gnmt}^j	Migration costs	—	Gravity and σ_N

Notes: Panel A reports calibrated and externally disciplined parameters. Panel B reports the dispersion of idiosyncratic tastes in the location and sector choices; larger values imply weaker responses to systematic payoff differences. Panel C reports kin parameters estimated from the death variation described in Section 5.2. Panel D lists fundamentals recovered by model inversion or auxiliary gravity equations. The parental connection is normalized to one in both kin margins. Horizontal relationship weights refer to the parent’s siblings.

5.2 Estimating the kin parameters

The kin block contains two sets of parameters: the value of geographically proximate kin in location choice and the value of local sector-specific kin in sector choice. The parameters are

$$\Theta^F = \{\psi_R^H, \Theta_H, \psi_R^A, \psi_R^M, \Lambda_A, \Lambda_M\}, \quad \Theta_H \equiv \frac{\theta_H}{\sigma_N}, \quad \Lambda_k \equiv \frac{\theta_k}{\sigma_O}.$$

The parent is the reference connection and is normalized to contribute one unit. The parameters ψ_R^H and ψ_R^k measure the contribution of an aunt-or-uncle connection relative to a parent. The

parameters Θ_H and Λ_k measure the value of a parental connection relative to the dispersion in idiosyncratic location and sector preferences. Once σ_N and σ_O are estimated, I recover the corresponding utility parameters θ_H and θ_k .

To identify these parameters, I exploit deaths of kin that remove particular connections. I use three linked twenty-year panels: 1900–1920, 1910–1930 and 1920–1940. For a panel beginning in t , eligible deaths occur between t and $t + 10$ and choices are measured in $t + 20$. Let $a = (g, m, j)$ index an ancestral-type-location-sector cell. Within each cell, I construct the pre-death population-weighted mean number of relevant connections lost and the corresponding choice log odds.³¹

Step 1: geographic kin attachment. I identify the value of geographically proximate kin using deaths of co-located *non-employed* parents and aunts or uncles. Restricting the deceased to non-employed relatives isolates the loss of local kin support from any contemporaneous change in sector-specific kin employment ties. In the model, a parent contributes one unit to the kin amenity at the origin. An aunt or uncle contributes ψ_R^H , weighted by the probability that the horizontal connection is inherited through parental sector q . Let $\pi_{gnjt}(q) \equiv \Pr_{t-1}(q \mid n, j, g)$ denote this probability.

To see how these deaths identify the location parameters, define $\Delta H_{gnjt}(m) \equiv H_{gnjt}(m) - H_{gnjt}(n)$ and $\Delta O_{gnjt}(m) \equiv O_t(m, j, g) - O_t(n, j, g)$. The log odds of leaving the baseline region n rather than remaining there are then

$$\ell_{gnjt}^N \equiv \log \frac{\sum_{m \neq n} P_{t+1}^N(m \mid n, j, g)}{P_{t+1}^N(n \mid n, j, g)} = \log \sum_{m \neq n} \exp \left\{ \kappa_{gnt}^j(m) + \Theta_H \Delta H_{gnjt}(m) + \frac{\beta}{\sigma_N} \Delta O_{gnjt}(m) \right\}. \quad (26)$$

where $\Theta_H = \theta_H / \sigma_N$ and $\kappa_{gnt}^j(m)$ collects current non-kin differences in the payoff from living in destination m rather than remaining in n , including wages, amenities, congestion and migration costs. The final term captures differences in children’s continuation opportunities across locations.

A co-located parent death lowers $H_{gnjt}(n)$ by one while leaving kin support at alternative destinations unchanged. An aunt-or-uncle death inherited through parental sector q lowers origin support by $\psi_R^H \pi_{gnjt}(q)$. Importantly, neither death mechanically changes O_t : the deceased parent or aunt or uncle of the current adult is a grandparent or more distant relative from the perspective of the next generation and therefore does not enter its kin connections by Assumption 2. Restricting the deceased to non-employed relatives further ensures that the death does not simultaneously remove a sector-specific employment connection.

³¹Using the same pre-death weights for treatment and outcomes preserves the linear model mapping because the structural coefficients are common across cells. The event studies in Section 3 instead report individual-level effects in percentage points. The regressions here use model-consistent choice log odds, so the two sets of coefficients are not numerically identical.

Let P denote a parent death and R an aunt-or-uncle death. The model therefore implies

$$\ell_P^{N,\text{post}} - \ell_P^{N,\text{pre}} = \Theta_H + \xi_P^N, \quad \ell_R^{N,\text{post}} - \ell_R^{N,\text{pre}} = \Theta_H \psi_R^H \pi_{gnjt}(q) + \xi_R^N, \quad (27)$$

where ξ_P^N and ξ_R^N collect changes in non-kin location payoffs and continuation values not mechanically generated by the loss of kin support. Conditional on controls and fixed effects, the identifying assumption is that these residual changes are unrelated to the timing of the death. The parent coefficient therefore identifies Θ_H , while the aunt-or-uncle coefficient relative to the parent coefficient identifies ψ_R^H .

Guided by this mapping, I estimate

$$\bar{\ell}_{a,t+20}^N = \alpha_a + \delta_{njt} + \beta_P^H \overline{\text{Dth}}_{at}^{P,H} + \beta_R^H \overline{\text{Dth}}_{at}^{R,H,\pi} + X'_{at} \Gamma_N + \varepsilon_{at}^H, \quad (28)$$

where $\overline{\text{Dth}}_{at}^{P,H}$ is the mean number of co-located non-employed parental connections lost and $\overline{\text{Dth}}_{at}^{R,H,\pi} \equiv \pi_{gnjt}(q) \overline{\text{Dth}}_{at}^{R,H}$ is the corresponding aunt-or-uncle mass weighted by the parental-sector probability implied by the model. Parent deaths require no weighting because the parental connection enters the location amenity with unit weight. The fixed effects δ_{njt} compare individuals with the same origin, own sector, parental sector and period, while X_{at} contains predetermined kin composition and local characteristics.

The remaining identification requirement is an exclusion restriction on the timing of deaths. Conditional on the fixed effects and controls, residual variation in parent and aunt-or-uncle deaths must be unrelated to changes in non-kin determinants of subsequent migration:

$$\mathbb{E}[\widetilde{\text{Dth}}_{at}^{P,H} \varepsilon_{at}^H] = 0, \quad \mathbb{E}[\widetilde{\text{Dth}}_{at}^{R,H,\pi} \varepsilon_{at}^H] = 0.$$

The assumption requires that, after conditioning on the initial state and predetermined characteristics, the timing of a death is unrelated to shocks that would independently change the individual's incentive to migrate. Origin-by-own-sector-by-parental-sector-by-year fixed effects absorb shocks common to workers in the same initial local state, while ancestry fixed effects and predetermined controls absorb persistent differences in kin composition and mortality exposure. Restricting the deceased to non-employed relatives additionally rules out the direct sector-employment channel described above. Under this exclusion restriction,

$$\hat{\Theta}_H = \hat{\beta}_P^H, \quad \hat{\psi}_R^H = \frac{\hat{\beta}_R^H}{\hat{\beta}_P^H}. \quad (29)$$

The estimates imply $\hat{\Theta}_H = 0.024$ and $\hat{\psi}_R^H = 0.874$. Thus, horizontal kin receive substantial weight in location choice relative to the normalized parental connection.

Table 3: Kin effects in location and sector choice

	Location choice	Sector choice	
	Home support (1)	Agriculture (2)	Manufacturing (3)
Panel A. Death coefficients			
Parent	0.024*** (0.002)	0.0206*** (0.003)	0.0489*** (0.012)
Aunt/uncle	0.021*** (0.002)	0.006*** (0.001)	0.007*** (0.003)
Panel B. Implied relative kin weights			
Aunt/uncle relative to parent, $\hat{\psi}_R$	0.874	0.305	0.141
Observations	962,098	611,273	611,273

Notes: Column (1) reports the parent and aunt-or-uncle coefficients from equation (28). Columns (2)–(3) report the corresponding sector-specific estimates from equation (30). Panel B reports the implied weight on one aunt-or-uncle connection relative to one parental connection, computed as $\hat{\psi}_R^H = \hat{\beta}_R^H / \hat{\beta}_P^H$ in column (1) and $\hat{\psi}_R^k = \hat{\beta}_R^k / \hat{\beta}_P^k$ in columns (2)–(3). Ratios are calculated using the unrounded coefficient estimates. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Step 2: local kin connections and sector choice. I next use deaths of co-located sector-employed parents and aunts or uncles to identify the kin-sector parameters. This restriction follows equation (8): sector access depends on kin who live in the same region and work in the relevant sector. In sector k , an employed parent contributes one unit to F^k , while each employed aunt or uncle contributes ψ_R^k .

Let $\ell \neq k$ denote the alternative sector. The log odds of choosing sector k rather than ℓ are

$$\ell_{gnjt}^k = \Lambda_k F_{gnjt}^k - \Lambda_\ell F_{gnjt}^\ell - \frac{\tau_{gnjt}^k - \tau_{gnjt}^\ell}{\sigma_O} + \frac{V_{t+1}(n, k, g) - V_{t+1}(n, \ell, g)}{\sigma_O},$$

where $\Lambda_h = \theta_h / \sigma_O$. The first two terms capture sector-specific kin support, the third other differences in sector-entry costs and the final term the value of the subsequent location choice. Consider the death of a co-located relative employed in sector k . A parent death reduces F^k by one, while an aunt-or-uncle death reduces it by ψ_R^k . The direct effect therefore lowers the log odds of choosing k by Λ_k and $\Lambda_k \psi_R^k$, respectively. There is one additional mechanical effect. Because sector choice precedes the individual's subsequent location choice, a co-located sector-employed relative also provides geographic support at the origin. Its death can therefore change the continuation value of choosing sector k relative to ℓ .

Let Ω_r^k denote this continuation-value response for a death of relationship type $r \in \{P, R\}$. The

model-implied change in sector- k log odds is

$$\ell^{k,\text{post}(r)} - \ell^{k,\text{pre}} = -\Lambda_k d_{F,r}^k + \Omega_r^k, \quad d_{F,P}^k = 1, \quad d_{F,R}^k = \psi_R^k.$$

The first term captures the loss of sector-specific kin support and the second the induced change in the value of the subsequent location choice. Without the latter, the parent response identifies Λ_k , while the aunt-or-uncle response relative to the parent identifies ψ_R^k . Appendix A4.4 derives Ω_r^k from the location problem. The correction depends on the difference in the probability of remaining at the origin after choosing sectors k and ℓ . If location behavior is identical across sectors, it cancels exactly.

I estimate

$$\bar{\ell}_{a,t+20}^k = \alpha_a^k + \delta_{njt}^k - \beta_P^k \overline{\text{Dth}}_{at}^{P,k} - \beta_R^k \overline{\text{Dth}}_{at}^{R,k} + X'_{at} \Gamma_k + \varepsilon_{at}^k, \quad (30)$$

where the death variables measure the mean numbers of co-located sector- k parental and aunt-or-uncle connections lost during the first decade of the panel. The fixed effects absorb common shocks within the initial local and sectoral state, while X_{at} contains predetermined kin composition, connections to the other sector, local characteristics and lagged sector-specific migration rates.

Identification requires that, conditional on these fixed effects and controls, death timing be unrelated to shocks that independently change the relative attractiveness of sectors k and ℓ . The age restrictions focus on premature deaths, while the pre-event evidence in the empirical section and the comparison with deaths of non-employed relatives address potential alternative explanations based on pre-existing trends or the consequences of losing a relative unrelated to sector-specific employment connections.

Under this exclusion restriction, the estimated coefficients satisfy

$$\beta_P^k = \Lambda_k - \bar{\Omega}_P^k, \quad \beta_R^k = \Lambda_k \psi_R^k - \bar{\Omega}_R^k,$$

where the bars denote continuation-value corrections averaged over the identifying variation. Appendix A4.4 shows that these corrections are quantitatively small, at most 2.8%. I therefore use the direct mapping in the baseline estimation:

$$\hat{\Lambda}_k = \hat{\beta}_P^k, \quad \hat{\psi}_R^k = \frac{\hat{\beta}_R^k}{\hat{\beta}_P^k}, \quad k \in \{A, M\}. \quad (31)$$

The resulting parameters are

$$\hat{\Lambda}_A = 0.0206, \quad \hat{\Lambda}_M = 0.0489, \quad \hat{\psi}_R^A = 0.305, \quad \hat{\psi}_R^M = 0.141.$$

Thus, one aunt-or-uncle connection receives about 31 percent of the weight of a parent in

agriculture and 14 percent in manufacturing. Appendix A4.4 shows that explicitly accounting for the continuation-value adjustment changes these parameters only slightly.

5.3 Migration costs and location values

I construct migration distance using the least-cost path between CZ centroids on the historical railroad network from [Atack \(2016\)](#), allowing travel off the rail network to be more costly. The construction adapts the least-cost network approach of [Donaldson and Hornbeck \(2016\)](#) to obtain a reduced-form proxy for passenger migration costs.³²

Let $\tilde{d}_{nmt}$ denote the log of rail-adjusted least-cost distance between origin n and destination m , with the within-CZ value normalized to zero. Let F_{gnt}^j denote the migration flow of ancestral type g and sector j from n to m . The location-choice equation implies

$$\mathbb{E}[F_{gnt}^j] = \exp\left(\text{orig}_{gnt}^j + \text{dest}_{gnt}^j + \hat{\Theta}_H H_{gnt}(m) - \lambda_t^j \tilde{d}_{nmt}\right), \quad (32)$$

where, within each year-sector regression, the origin fixed effect absorbs origin mass and the location-choice denominator, while the destination fixed effect absorbs components of destination utility common across origins. Because $H_{gnt}(m)$ varies jointly with origin and destination, it is not absorbed by these fixed effects. I therefore estimate equation (32) by PPML ([Silva & Tenreyro, 2006](#)), treating the previously estimated $\hat{\Theta}_H H_{gnt}(m)$ as known. The coefficient λ_t^j then identifies the distance-related component of migration costs.

The distance coefficient identifies migration costs relative to the location taste dispersion:

$$\frac{M_{gnt}^j}{\sigma_N} = \lambda_t^j \tilde{d}_{nmt}. \quad (33)$$

After estimating σ_N , migration costs in utility units are $M_{gnt}^j = \sigma_N \lambda_t^j \tilde{d}_{nmt}$.

The estimated destination values, together with the kin and distance terms, imply the location inclusive value

$$\frac{V_t(n, j, g)}{\sigma_N} = \log \sum_{m \in N} \exp\left(\text{dest}_{gnt}^j + \hat{\Theta}_H H_{gnt}(m) - \hat{\lambda}_t^j \tilde{d}_{nmt}\right). \quad (34)$$

Railroad placement is not random. Lines may have been built along economically important corridors that also attracted migration and shaped where families settled. Origin and destination fixed effects absorb differences associated with the two endpoints, but not unobserved bilateral attractiveness correlated with rail connectivity. I therefore interpret $\hat{\lambda}_t^j \tilde{d}_{nmt}$ as a reduced-form measure of bilateral migration costs rather than the causal effect of railroad access. This concern does not identify the kin coefficient Θ_H , which is estimated separately

³²Appendix A4.3 provides the details and sensitivity to the off-rail cost multiplier.

from the death variation above, but it can affect the interpretation of the recovered distance component.

5.4 Indirect inference

I next estimate the dispersion of idiosyncratic location and sector tastes, (σ_N, σ_O) . These parameters determine how strongly choices respond to differences in systematic utility. A larger value implies more dispersed idiosyncratic tastes and therefore a weaker response of choices to a given change in wages, costs or amenities. They cannot, however, be recovered directly from observed migration or sectoral responses to wages. In particular, with non-homothetic preferences, a given change in wages or productivity does not translate into a constant change in utility, so there is no simple elasticity mapping from observed migration responses to σ_N (Eckert & Peters, 2022). I therefore use the Dust Bowl as an external productivity shock and choose (σ_N, σ_O) so that the full model reproduces its empirical migration and sectoral responses. The estimation follows the indirect-inference approach of Gouriéroux, Monfort, and Renault (1993), Bilal and Rossi-Hansberg (2023) and Rodríguez-Clare, Ulate, and Vásquez (2026).

The Dust Bowl provides a large and spatially heterogeneous agricultural productivity disturbance (Hornbeck, 2012, 2023; Liu, 2026). I construct two sets of twenty-year response moments. For sector choice, I follow children younger than 15 at baseline and measure manufacturing employment twenty years later, separately by parental sector. For migration, I follow adults older than 25 and measure whether they leave their baseline CZ, separately by initial sector. In both cases, I estimate the differential responses of medium- and high-erosion CZs relative to low-erosion CZs using CZ and state-by-year fixed effects and predetermined controls. Appendix A4.5 reports the specifications and samples.

I separately recover the size of the underlying agricultural productivity disturbance from the static goods-market inversion, which does not depend on (σ_N, σ_O) . By 1950, recovered agricultural productivity is 0.179 log points lower in medium-erosion CZs and 0.226 log points lower in high-erosion CZs, with no comparable pre-Dust-Bowl differences. As an external check, farm products per acre fall by 0.176 and 0.241 log points in the same two groups. The similarity of these magnitudes is not imposed by the inversion and supports the size of the productivity disturbance used for identification.

For each candidate $\sigma = (\sigma_N, \sigma_O)$, I first combine the kin parameters identified relative to taste dispersion with each candidate (σ_N, σ_O) to recover their utility-unit counterparts and re-invert the model so that the candidate parameterization rationalizes the observed pre-counterfactual choices. I then solve two equilibrium transitions. The first follows the observed economy, including the estimated Dust Bowl productivity disturbance. The second removes only that component of agricultural productivity while holding the remaining recovered fundamentals fixed. Sector choices, migration, wages, prices and kin connections are allowed to adjust en-

dogenously in both economies.

Applying the same auxiliary regressions used in the data to the simulated outcomes gives a vector of model-implied responses $b^{\text{model}}(\sigma)$. I choose

$$\hat{\sigma} = \arg \min_{\sigma \in \Sigma} \left[b^{\text{model}}(\sigma) - \hat{b}^{\text{data}} \right]' W \left[b^{\text{model}}(\sigma) - \hat{b}^{\text{data}} \right], \quad W = I. \quad (35)$$

Thus, the dispersions are identified by the response of migration and sector choice to an independently measured productivity shock, while the kin parameters are disciplined separately by the death variation above. Appendix A4.5 provides the full construction, including the candidate-specific inversion, the simulated auxiliary regressions and the maintained identifying assumptions.

Estimation results and model fit. The estimation yields $(\hat{\sigma}_N, \hat{\sigma}_O) = (0.51, 0.60)$. These estimates are of a similar order of magnitude to dispersion parameters used in related quantitative spatial and sectoral-choice models (Eckert & Kleineberg, 2024; Eckert & Peters, 2022; Rodríguez-Clare et al., 2026). At the selected values, I perform the final inversion and use the resulting fundamentals and kin parameters in the quantitative exercises that follow.

Appendix A4.6.3 reports several untargeted checks on the resulting economy. Recovered agricultural and manufacturing productivity covary with independent measures of sectoral value per worker that are not used in the inversion. The model also reproduces the spatial pattern of employment growth from 1880 to 1950, including the shift from agriculture toward manufacturing in initially agricultural places. Finally, the kin block matches two features not targeted in estimation: the cross-CZ distribution of aunts and uncles and the fact that most aggregate structural change occurred through sectoral reallocation within commuting zones rather than population reallocation across them.

6 Quantitative importance of kinship

The quantitative analysis uses the model to assess the economic importance of the kin channels identified in the empirical analysis. I begin by expressing the value of kin connections relative to the other sectoral and spatial barriers workers face. I then quantify their aggregate consequences and investigate their importance in spatial policies.

6.1 Kinship and barriers to reallocation

One explanation for persistent wage gaps across sectors is that workers face barriers to moving across occupations and places (Donovan & Schoellman, 2023). Since these barriers are not directly observed, quantitative models recover them as residual sectoral and spatial wedges

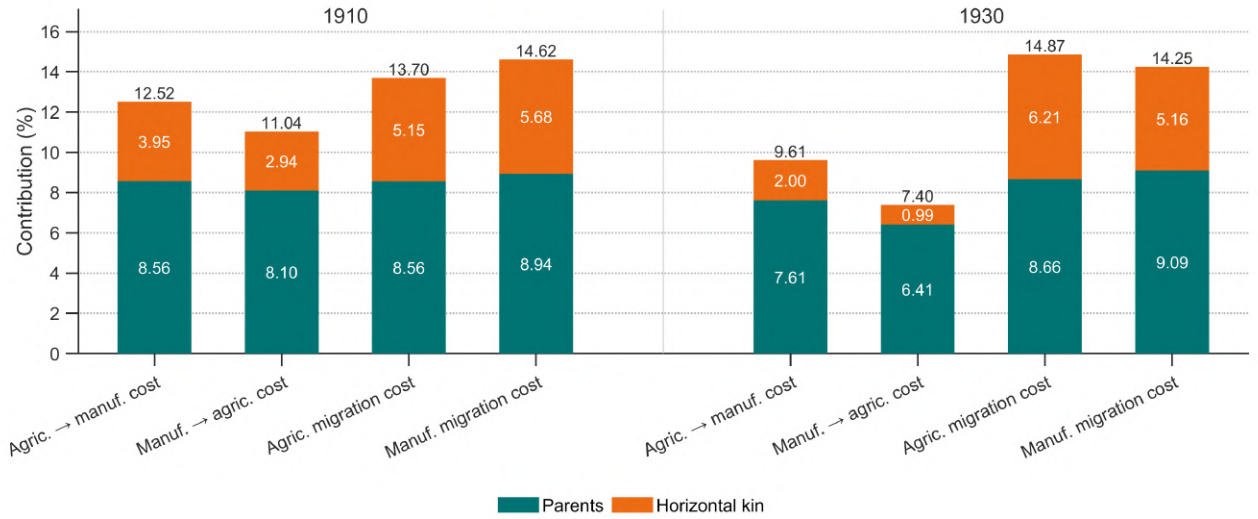
that rationalize observed choices (Eckert & Kleineberg, 2024; Takeda, 2022). Kinship and social ties have long been viewed as one source of such barriers because they can keep workers attached to their current sector and place. The linked kin data and model structure allow me to quantify this channel directly. I first measure how much sector- and location-specific kin connections contribute to the recovered barriers. I then ask how these connections affect structural transformation.

To quantify how much these channels contribute to the recovered barriers, I remove each kin channel in turn and re-invert the model while holding observed sector and location choices fixed. The corresponding residual wedge must then adjust to compensate for the missing kin channel. For recovered wedge X , I compare its counterfactual value $X_{it}^{(-c)}$, obtained after removing kin channel c , with its baseline value X_{it} . I report

$$\mathcal{L}_{ct}^X = 100 \frac{\sum_i \ell_{it} |X_{it}^{(-c)} - X_{it}|}{\sum_i \ell_{it} |X_{it}|}, \quad (36)$$

where ℓ_{it} is employment in kin-origin-sector group i . The statistic measures the employment-weighted change in the recovered wedge required to reproduce the same observed sector and location choices when kin channel c is removed.

Figure 6: Kinship and barriers to reallocation



Notes: The figure reports the kin contribution to recovered sector-entry and migration barriers in 1910 and 1930. Bars show the employment-weighted re-inversion adjustment relative to the recovered object, $\mathcal{L}$, defined in equation (36). Each bar decomposes the adjustment into parental and horizontal-kin components.

Figure 6 shows that kinship accounts for a meaningful portion of both sectoral and spatial barriers. In 1910, removing kin channels requires an adjustment equal to 12.52 percent of the recovered agriculture-to-manufacturing entry cost and 11.04 percent of the manufacturing-

to-agriculture cost. The corresponding adjustments to migration costs are 13.70 percent for agricultural workers and 14.62 percent for manufacturing workers. Horizontal kin are quantitatively important alongside parents: they account for 32 percent of the kin component of the agriculture-to-manufacturing barrier.

The composition of this contribution changes over development. Between 1910 and 1930, the kin component of the agriculture-to-manufacturing barrier falls from 12.52 to 9.61 percent and the manufacturing-to-agriculture component falls from 11.04 to 7.40 percent. The contribution to migration barriers instead rises to 14.87 percent for agricultural workers, while remaining broadly stable at 14.25 percent for manufacturing workers. Kinship therefore remains quantitatively important as structural transformation proceeds, but its incidence shifts from sectoral entry toward spatial adjustment.

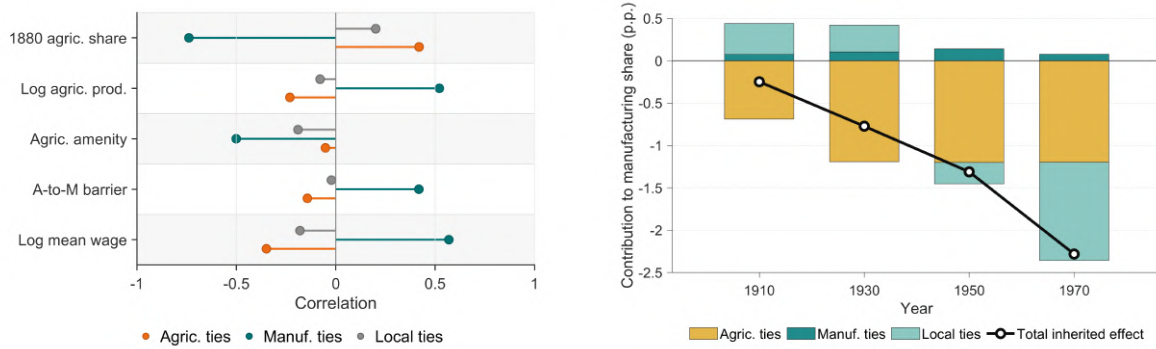
The initial distribution of kin. The effect of kin on development depends not only on the size of the barrier but also on where relatives live and which sectors they work in. Panel 7a shows that these connections are systematically related to the economic conditions families face in 1880. Agricultural and local connections are more common among families originating in relatively agricultural and lower-opportunity places, while manufacturing connections are more common where manufacturing is already more attractive. The initial distribution of kin therefore may reinforce the sectoral and spatial allocation at the start of the transition.

The aggregate effects of the initial kin distribution. I next isolate the effect of the kin distribution inherited at the start of structural transformation. To do so, I compare an economy without kinship with an otherwise identical economy in which kinship is present but their sectoral and spatial distribution is constant at its observed 1880 pattern. This asks how kinship initiated at the beginning of the transition affected the movement out of agriculture, before allowing migration and sector switching to change where relatives lived and worked.

Panel 7b shows that the 1880 distribution of kin becomes a larger drag on structural transformation as the economy moves away from its initial allocation. By 1950, holding kinship constant at their 1880 distribution lowers the manufacturing share by 1.311 percentage points relative to an economy without kinship. This amounts to around 1.8 percent of the 1950 manufacturing share. Agricultural connections account for most of the decline, reducing the manufacturing share by 1.200 points. Manufacturing connections offset 0.142 points, while local connections reduce the share by another 0.253 points. By 1970, the total effect reaches -2.281 percentage points, equivalent to 4.1 percent of the increase in manufacturing between 1880 and 1970.

The effect grows over time. As manufacturing expands, connections to agriculture increasingly reinforce a declining sector, while local connections continue to favor the places in which families were concentrated in 1880. The same inherited kin distribution therefore becomes more restrictive as economic activity shifts across sectors and space.

Figure 7: The initial kin distribution and structural transformation



(a) Allocation of kin in 1880

(b) Effect of the 1880 kin distribution

Notes: Panel (a) reports CZ-level correlations in 1880 between agricultural, manufacturing and local kin connections and selected local characteristics inverted from the model. Panel (b) reports contributions to the national manufacturing share in percentage points from holding the observed 1880 kin distribution constant relative to an economy without kinship. Bars report the exact Shapley decomposition into agricultural, manufacturing and local connections. The black line reports the total effect of the 1880 kin distribution.

Where those relatives are located is central to this result. Appendix A5.3 keeps the same total kin mass observed in 1880 but redistributes relatives across CZs in proportion to population. Under this population-proportional distribution, kin lower the 1950 manufacturing share by only 0.03 percentage points. The observed geography accounts for the remaining 1.28 points of the 1.31-point effect. The initial drag therefore comes primarily from the spatial distribution of relatives rather than from the amount of kin alone.

The effect operates through both sectoral change within places and the spatial distribution of employment. In 1950, 0.688 percentage points of the 1.311-point decline come from lower manufacturing shares within CZs and 0.603 points from changes in where employment is located, with only a small interaction term. The initial kin distribution therefore slows structural transformation both by affecting sector choice within places and by affecting movement across places.

These results show why kinship can act as a barrier to reallocation. Workers inherit connections to particular sectors and places and those connections become more or less valuable as economic opportunities change. The analysis so far has held the 1880 distribution of kin constant. In practice, migration and sector switching change where relatives live and work, which in turn changes the connections inherited by later generations. I next quantify how this evolution of kin affects the pace of structural transformation.

6.2 Kinship propagation and structural transformation

The previous subsection held the 1880 distribution of kin constant to isolate the incidence of relatives. I now allow migration and sector switching to change where relatives live and which sectors they work in. I compare the full model with an otherwise identical economy in which kin remain constant at their observed 1880 sectoral and spatial distribution while all non-kin fundamentals evolve identically in both economies. The difference therefore measures how changes in kin generated by workers' own reallocation affect subsequent structural transformation.

Evolving kin connections accelerate structural transformation. The previous results show that the kin distribution in 1880 slows reallocation. This need not remain true as structural transformation proceeds, however, because the same migration and sector switching that drive the transition also change where relatives live and work. I therefore ask how much this endogenous reorientation of kin offsets the initial drag.

Allowing kin to evolve raises the national manufacturing share by 0.727 percentage points in 1950 and by 2.053 points in 1970 relative to keeping their 1880 distribution constant. The latter accounts for 3.7 percent of the entire increase in the manufacturing share between 1880 and 1970. This propagation advances structural transformation by approximately seven years: the economy reaches the manufacturing share of the constant-kin economy in 1970 already by 1963. These changes offset 55 percent of the initial drag from the 1880 distribution by 1950 and approximately 90 percent by 1970. By the end of the period, the net effect of kin on the national manufacturing share is therefore only -0.228 percentage points.

The additional manufacturing is concentrated in places that begin the transition further behind. Table 4 shows that, by 1970, allowing kin to evolve raises the manufacturing share in the South by 3.1 percentage points and in CZs that were in the bottom third of the 1880 manufacturing distribution by 3.8 points. The share of the population living in a manufacturing-majority CZ rises by 1.9 percentage points. Relative to the overall transition, these effects are especially important for spatial convergence. Changing kin connections explain about 4 percent of the national rise in manufacturing, but 17 percent of the narrowing of the manufacturing gap between the most and least industrialized thirds of CZs and 19 percent of the decline in the dispersion of manufacturing shares across CZs. They also account for 8 percent of the across-CZ component of structural change, compared with 3 percent of the within-CZ component. Changing kin connections therefore contribute disproportionately to the geographic diffusion of industrialization and the catch-up of places that begin the transition behind (Eckert & Peters, 2022).

The source of the effect changes over development. Panel 8a separates the manufacturing effect into four channels. Changes in horizontal kin can operate through the sector and

Table 4: Effects of evolving kin connections on structural transformation

<i>Panel A. Aggregate structural transformation</i>			
	Full	Constant 1880	Effect
Manufacturing share in 1970 (%)	79.7	77.6	+2.1 pp
Year reaching constant-kin 1970 manufacturing share	1963	1970	7 years earlier
<i>Panel B. Spatial incidence</i>			
	Full	Constant 1880	Effect
Population in manufacturing-majority CZs (%)	90.8	88.9	+1.9 pp
Manufacturing share in the South (%)	71.3	68.2	+3.1 pp
Manufacturing share in bottom-third CZs (%)	62.9	59.1	+3.8 pp
<i>Panel C. Contribution to structural change and convergence</i>			
	Full-model change	Share explained by kin	
Manufacturing share (pp)	55.1	4%	
Within CZs (pp)	48.7	3%	
Across CZs (pp)	6.4	8%	
Manufacturing share, bottom-third CZs (pp)	60.1	6%	
Top–bottom third manufacturing-share gap (pp)	−16.8	17%	
SD of CZ manufacturing shares (pp)	−6.5	19%	

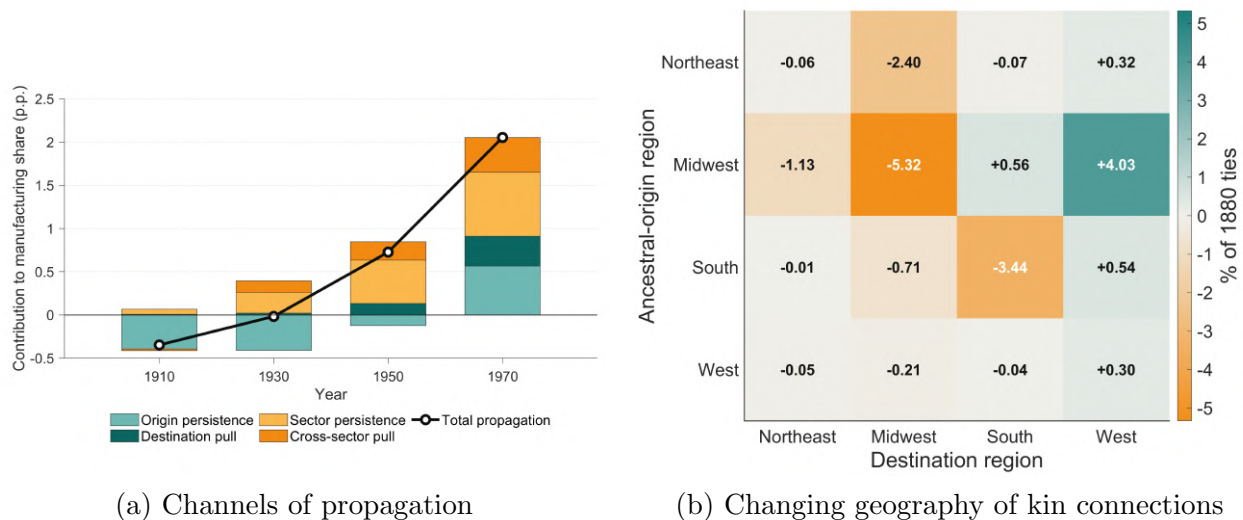
Notes: The full model allows kin connections to evolve endogenously with sector and location choices. The constant-kin counterfactual holds the sectoral and spatial distribution of non-parental relatives at its observed 1880 pattern, while parental connections and all non-kin fundamentals evolve as in the full model. Panels A and B report outcomes in the two economies and their difference. Panel C reports the share of the 1880–1970 change in each outcome explained by evolving kin connections, calculated as the difference between the full and constant-kin economies relative to the change in the full model. Thirds of CZs are defined by their manufacturing share in 1880. The within- and across-CZ components come from a shift-share decomposition. The standard deviation of CZ manufacturing shares is weighted by employment. For the timing measure, the full-model path is linearly interpolated between model periods.

location inherited from the preceding generation or through connections to alternative sectors and destinations. Panel 8b shows how the geography of horizontal kin changes across broad regions.

Sectoral changes matter earlier than spatial changes. In 1950, sector persistence and cross-sector pull account for 0.718 of the 0.727-percentage-point manufacturing effect, while origin persistence and destination pull nearly offset one another. By 1970, the sectoral channels contribute 1.142 percentage points and the spatial channels 0.911, so changes in the geography of kin account for 44 percent of the long-run effect. The spatial channel takes longer to emerge because migration must first change where relatives live before those new locations affect the choices of later generations.

The decomposition also separates access to new opportunities from the persistence of earlier

Figure 8: Channels and geography of horizontal-kin propagation



Notes: Panel (a) reports the Shapley decomposition of the manufacturing-share effect of changing kin into origin persistence, destination pull, sector persistence and cross-sector pull. Origin and destination refer to successive generations. The black line reports the total effect. Panel (b) reports the net change between 1880 and 1970 in population-weighted kin connections for each ordered pair of Census regions, excluding connections at the ancestral CZ itself, as a percentage of the 1880 stock. Rows denote ancestral-origin regions and columns destination regions.

reallocation. By 1970, destination pull and cross-sector pull contribute a combined 0.707 percentage points, about one third of the total effect. Origin persistence and sector persistence contribute 1.346 points, about two thirds. Most of the long-run effect therefore comes not from repeatedly drawing workers away from the same original sector and place, but from reinforcing the sectors and places reached by earlier generations.

The spatial channel makes this mechanism particularly clear. Origin persistence is negative in 1950 but positive by 1970 because the relevant origin changes across generations. Once a family branch migrates, a destination for one generation becomes the origin of the next. The same mechanism operates across sectors: once a branch enters manufacturing, later relatives inherit more connections to manufacturing. Kinship therefore increasingly reinforces the new allocation created by earlier migration and sector switching compared to the allocation observed in 1880.

Two additional exercises in Appendix A5.4 clarify how this process accumulates. About three quarters of the 1970 manufacturing effect comes from higher manufacturing shares within CZs, while the remainder comes from changes in the spatial distribution of employment. The effect also requires repeated rounds of adjustment. Allowing kin to evolve only through 1910 generates essentially none of the eventual manufacturing gain, while allowing them to evolve through 1930 already produces 1.522 of the eventual 2.053 percentage points.

The composition of kin changes accordingly. Agriculture persistence falls from 0.452 units of

kin mass in 1880 to 0.223 in 1970, while manufacturing persistence rises from 0.116 to 0.463. By 1970, 86.6 percent of newly created destination connections are supplied by relatives working in manufacturing. Kin that initially reinforce agriculture and inherited origins therefore become increasingly connected to manufacturing and to new places.

6.3 Spatial policies and kinship spillovers

The model shows that workers' choices create spillovers to other family branches that they do not take into account. Moving gives relatives a connection to a new destination, while entering manufacturing gives them a connection to that sector. These spillovers create a role for relocation policy, particularly early in structural transformation when kin are concentrated in agriculture and less industrialized places.

These spillovers make kinship relevant for a spatial-policy question that matters today as much as it did in the past: should policy bring economic activity to workers in declining places or help workers move toward expanding opportunities? During the New Deal, a 1941 study by the Works Progress Administration (WPA) of depressed coal towns, for example, contrasted attracting new industry with “guided migration” toward expanding areas.³³ I study relocation assistance to examine how kinship shapes migration and sectoral reallocation jointly, including the spillovers to other relatives and later generations.

To quantify these forces, I introduce a one-time (generation) moving voucher in 1880 for workers leaving targeted low-manufacturing CZs. I set the benchmark budget to one percent of the 1880 national wage bill, a sizable but plausible policy scale given the large resources devoted to historical relocation programs such as Indonesia's Transmigration Program and China's Anti-Poverty Relocation Program (Bazzi et al., 2016; Zhang et al., 2023).³⁴ I compare the full equilibrium with an economy in which the voucher has the same immediate effect on where workers live and work, but later generations inherit the kin connections they would have had without the voucher. The difference measures the additional effects generated through kinship. More specifically, the policy targets the 137 commuting zones in the bottom 30 percent of the 1880 manufacturing-share distribution. These places contain 20.9 percent of the population but only 2.25 percent of national manufacturing employment and are concentrated primarily in the South and lower Midwest (Appendix Figure A5.5.1). The voucher pays workers who leave a targeted CZ for a destination outside the targeted set. The one-percent budget amounts to

³³Brown and Webb's *Seven Stranded Coal Towns* (WPA, 1941) considered two responses directly: encouraging industrial expansion in depressed areas or “guided migration” of surplus workers toward active areas, supported by training, job placement and travel assistance.

³⁴The comparison is intended to discipline the order of magnitude rather than match the fiscal scale of a particular program. Indonesia's Transmigration Program relocated about two million people between 1979 and 1988, while China's Anti-Poverty Relocation Program relocated more than 9.6 million people between 2016 and 2020. Results are qualitatively similar for budgets between 0.25% and 10% of the 1880 wage bill.

\$119.4 million and is financed by a uniform lump-sum tax, implying a relocation grant of \$723, or 20.8 percent of the average annual wage.

Kinship is a quantitatively important component of moving costs. I first ask how much workers value the kinship ties they give up when they move. For each origin–destination pair, I calculate the transfer that would compensate a worker for the difference in kinship ties between the origin and destination. Figure 9 reports the distribution of these compensating transfers. The median is \$157, equal to approximately 12 percent of annual earnings and 22 percent of the \$723 relocation grant in the benchmark counterfactual. The 75th and 90th percentiles are \$249 and \$353.

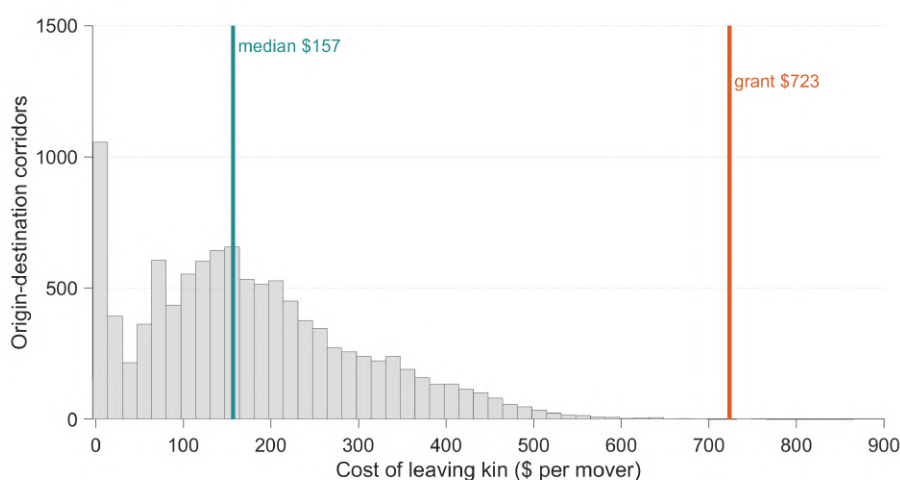


Figure 9: The value of kinship ties lost through relocation

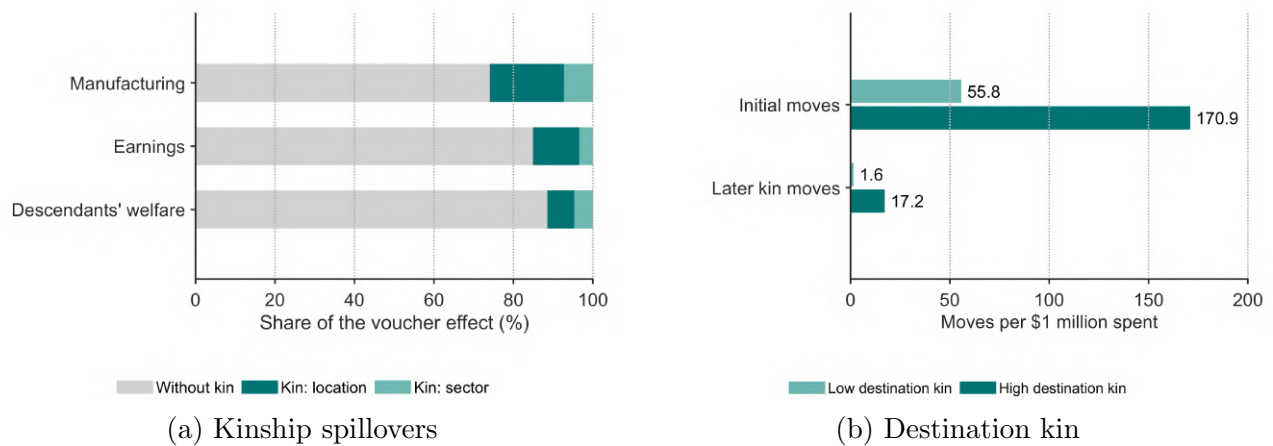
Notes: The histogram reports, across targeted origin–destination pairs, the transfer that compensates a worker for the difference between kinship ties at the origin and at the destination. The teal line marks the median compensating transfer of \$157. The orange line marks the \$723 relocation grant under the benchmark one-percent budget.

Quantifying spillovers through kinship. The relocation voucher can affect migration beyond the initially treated generation. To measure how much subsequent migration follows from the initial policy response, I report the change in later moves per 100 workers induced to move in 1880. To isolate the additional effects through kinship, I compare the full equilibrium with an economy in which the voucher has the same immediate effect on where workers live and work, but later generations inherit the kin connections they would have had without the policy. Under this counterfactual, cumulative migration between 1910 and 1970 falls by 34.1 moves per 100 initial movers relative to no voucher. Allowing the initial moves to reshape kin connections offsets about two fifths of this decline, generating 13.9 additional later moves per 100 initial movers. The voucher therefore does more than shift migration toward the first generation: by

placing relatives in new destinations, it makes those destinations easier for other kin to reach later.

The same spillover materially changes the policy’s longer-run economic effects. Figure 10 separates the direct effect of the voucher from the additional effect that arises because the initial moves change the kin connections inherited by later generations, distinguishing the locations and sectors of extended kin.

Figure 10: Kinship and the effectiveness of moving vouchers



Notes: Moving voucher with a budget equal to one percent of the 1880 wage bill. Panel (a) normalizes the total policy effect to 100 and decomposes it into the effect that remains when initial moves do not change later kin connections and the additional contributions operating through kin. Manufacturing and earnings are measured in 1970; welfare aggregates later generations from 1930 to 1970. Panel (b) reports initial and subsequent kin moves per \$1 million of policy expenditure when eligible destinations have relatively high or low levels of destination kin. Subsequent kin moves are measured through 1950.

Kinship spillovers account for a substantial share of the policy’s long-run effects. Ignoring the effect of an initial mover on other kin understates the voucher’s 1970 effect on manufacturing employment by 26 percent, its earnings effect by 15 percent and the earnings-based welfare gain of later generations by approximately 12 percent.

The location of extended kin is the largest channel. When a worker moves, other kin gain a connection to a destination they may not previously have been able to access through relatives. This lowers the kin-related cost of moving there in later generations. Sector switching creates a parallel channel: when a worker enters manufacturing, other kin gain a connection to that sector. Across all three outcomes, the destination channel is larger than the sectoral channel, consistent with the importance of destination kin for subsequent migration.

Destination kin amplify relocation assistance. I next ask whether having kin at potential destinations changes the effectiveness of relocation assistance. For each origin, I first restrict eligible destinations to the quarter with the highest predicted migration when extended-kin connections are excluded. I then divide these destinations according to the presence of initial

destination kin. The high-kin rule allows moves to the half with more destination kin, while the low-kin rule allows moves to the half with fewer. Both sets are therefore drawn from destinations that already rank highly on predicted migration without extended kin, but differ sharply in the number of relatives already living there.

The same fiscal resources generate very different initial responses. With a budget equal to one percent of the 1880 wage bill, the high-kin policy induces approximately three times as many movers as the low-kin policy (Figure 10, panel b). Equivalently, the fiscal cost per induced mover is about one third as large: \$5,852 compared with \$17,929. Existing destination kin therefore substantially increase the effectiveness of moving vouchers even before accounting for effects on later generations.

The difference becomes larger once spillovers to other kin are taken into account. Through 1950, the high-kin policy generates 10.1 additional later moves per 100 initial movers, compared with 2.9 under low-kin targeting. Kinship spillovers account for 46 percent of the high-kin policy's manufacturing effect but only 6 percent of the low-kin policy's effect. Destination kin therefore matter twice: they increase the initial response to the voucher and strengthen the effect of those moves on other kin.

7 Conclusion

This paper shows that kinship can both slow and propagate structural transformation. Relatives in current sectors and origins reinforce persistence, while relatives in expanding sectors and at destinations facilitate sector switching and migration. Because workers who switch sectors and move change where their relatives work and live, these effects build across generations. In the quantitative model, the initial concentration of kin in agriculture and less industrialized places slows structural transformation, but subsequent changes in kin connections offset 55% of this initial drag by 1950 and about 90% by 1970.

These dynamics have direct implications for relocation policy. A subsidized move does more than change the location of its recipient: it places a relative at a new destination and can therefore lower migration barriers for other kin and later generations. In a moving-voucher experiment, ignoring these spillovers understates the long-run manufacturing effect by 26%, the earnings effect by 15% and the welfare gains of later generations by about 12%. Existing destination kin also affect the cost of generating the initial response: among economically attractive destinations, the same budget induces about three times as many moves when more relatives already live there. Relocation policies can therefore be evaluated and targeted differently once kinship is taken into account. Evaluations that follow only direct recipients miss part of the long-run response, while targeting rules based only on economic conditions can overlook destinations where workers face substantially lower effective barriers to moving.

More broadly, the results imply that migration barriers are partly dynamic. A move today does not only reallocate one worker: it creates a destination connection that can make the same move easier for other relatives tomorrow. This generates increasing returns to relocation. Where destination kin already exist, a given subsidy induces more movement; where policy creates new destination kin, its effects can propagate across relatives and generations. Relocation policy can therefore have larger and more persistent effects than suggested by the response of its direct recipients alone.

Two extensions are particularly natural. First, marriage and household formation are taken as given, even though marriage combines family geographies and may itself respond to migration and occupational opportunities. Endogenizing family formation would allow structural change to reshape kin connections through an additional margin. Second, the analysis stops at the agriculture-to-manufacturing transition. The rise of services may place different weight on education, credentials and access to particular occupations, potentially changing both how kin matter and how quickly these connections adapt. Studying these margins would help establish how the role of kinship evolves across later stages of development.

References

- Adamopoulos, T., Brandt, L., Chen, C., Restuccia, D., & Wei, X. (2024). Land Security and Mobility Frictions. *Quarterly Journal of Economics*, 139(3), 1941–1987.
- Adermon, A., Lindahl, M., & Palme, M. (2021, May). Dynastic human capital, inequality, and intergenerational mobility. *American Economic Review*, 111(5), 1523–48. doi: 10.1257/aer.20190553
- Alesina, A., & Giuliano, P. (2014). Family ties. In P. Aghion & S. N. Durlauf (Eds.), *Handbook of economic growth* (Vol. 2, pp. 177–215). Elsevier. doi: 10.1016/B978-0-444-53538-2.00004-6
- Allen, T., & Donaldson, D. (2020). *Persistence and path dependence in the spatial economy* (NBER Working Paper No. 28059). National Bureau of Economic Research. doi: 10.3386/w28059
- Althoff, L., Brookes Gray, H. M., & Reichardt, H. (2026). *Mothers, schools, and the making of american human capital mobility* (Working Paper No. 35152). National Bureau of Economic Research. (Accepted at the Review of Economic Studies) doi: 10.3386/w35152
- Althoff, L., Gray, H. B., & Reichardt, H. (2025). *America’s rise in human capital mobility* (Working Paper).
- Antonie, L., Inwood, K., Minns, C., & Summerfield, F. (2020). Selection bias encountered in the systematic linking of historical census records. *Social Science History*, 44(3), 555–570. doi: 10.1017/ssh.2020.15
- Artuç, E., Chaudhuri, S., & McLaren, J. (2010). Trade shocks and labor adjustment: A structural empirical approach. *American Economic Review*, 100(3), 1008–1045. doi: 10.1257/aer.100.3.1008
- Attack, J. (2016, May). *Historical Geographic Information Systems (GIS) Database of U.S. Railroads for 1850, 1860, 1870, 1880, 1890, and 1900*. (Revised version.)
- Attack, J., Bateman, F., Haines, M., & Margo, R. A. (2010, January). *Did Railroads Induce or Follow Economic Growth? Urbanization and Population Growth in the American Midwest, 1850–1860* (Working Paper No. 14640). National Bureau of Economic Research. (Published in *Social Science History*, 35(2), 2011)
- Azoulay, P., Graff Zivin, J. S., & Wang, J. (2010). Superstar extinction. *The Quarterly Journal of Economics*, 125(2), 549–589. doi: 10.1162/qjec.2010.125.2.549
- Bailey, M., Cao, R., Kuchler, T., Stroebe, J., & Wong, A. (2018a). Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives*, 32(3), 259–280. doi: 10.1257/jep.32.3.259
- Bailey, M., Cao, R., Kuchler, T., Stroebe, J., & Wong, A. (2018b, August). Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives*, 32(3), 259–80.

- Bailey, M., Cole, C., & Massey, C. (2020). Simple strategies for improving inference with linked data: A case study of the 1850–1930 IPUMS linked representative historical samples. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 53(2), 80–93. doi: 10.1080/01615440.2019.1630343
- Barro, R. J., & Becker, G. S. (1989). Fertility Choice in a Model of Economic Growth. *Econometrica*, 57(2), 481–501. doi: 10.2307/1912563
- Baseler, T. (2025). Migration spillovers within families: Evidence from thailand. *Review of Economic Dynamics*, 55, 101255. doi: 10.1016/j.red.2024.101255
- Bau, N., & Fernández, R. (2021). *The Family as a Social Institution* (Working Paper No. 28918). National Bureau of Economic Research. doi: 10.3386/w28918
- Bazzi, S., Gaduh, A., Rothenberg, A. D., & Wong, M. (2016). Skill transferability, migration, and development: Evidence from population resettlement in indonesia. *American Economic Review*, 106(9), 2658–2698. doi: 10.1257/aer.20141781
- Bilal, A., & Rossi-Hansberg, E. (2023, June). *Anticipating climate change across the united states* (Working Paper No. 31323). National Bureau of Economic Research.
- Blumenstock, J. E., Chi, G., & Tan, X. (2025). Migration and the value of social networks. *The Review of Economic Studies*, 92(1), 97–128. doi: 10.1093/restud/rdad113
- Boppart, T. (2014). Structural change and the kaldor facts in a growth model with relative price effects and non-gorman preferences. *Econometrica*, 82(6), 2167–2196.
- Borusyak, K., Hull, P., & Jaravel, X. (2022). Quasi-experimental shift-share research designs. *Review of Economic Studies*, 89(1), 181–213. doi: 10.1093/restud/rdab030
- Büchel, K., Ehrlich, M. V., Puga, D., & Viladecans-Marsal, E. (2020). Calling from the outside: The role of networks in residential mobility. *Journal of Urban Economics*, 119, 103277. doi: 10.1016/j.jue.2020.103277
- Buckles, K., Haws, A., Price, J., & Wilbert, H. E. B. (2025). Breakthroughs in historical record linking using genealogy data: The census tree project. *Explorations in Economic History*, 98, 101717. doi: 10.1016/j.eeh.2025.101717
- Budí-Ors, T., & Pijoan-Mas, J. (2025). *Macroeconomic development, rural exodus, and uneven industrialization*. (Working paper, June 2025; revise and resubmit at JPE Macroeconomics)
- Caliendo, L., Dvorkin, M., & Parro, F. (2019). Trade and labor market dynamics: General equilibrium analysis of the china trade shock. *Econometrica*, 87(3), 741–835.
- Calvó-Armengol, A., & Jackson, M. O. (2004). The effects of social networks on employment and inequality. *American Economic Review*, 94(3), 426–454. doi: 10.1257/0002828041464542
- Cao, H., Chen, C., Xi, X., & Zuo, S. X. (2024). Family migration and structural transformation. *Canadian Journal of Economics*, 57(3), 753–776. doi: 10.1111/caje.12725

- Carrington, W. J., Detragiache, E., & Vishwanath, T. (1996). Migration with endogenous moving costs. *American Economic Review*, 86(4), 909–930. doi: 10.1257/aer.86.4.909
- Cassan, G., Keniston, D., & Kleineberg, T. (2021). *A Division of Laborers: Identity and Efficiency in India* (Working Paper No. 28462). National Bureau of Economic Research.
- Comin, D., Lashkari, D., & Mestieri, M. (2021). Structural change with long-run income and price effects. *Econometrica*, 89(1), 311–374.
- Compton, J., & Pollak, R. A. (2014). Family proximity, childcare, and women’s labor force attachment. *Journal of Urban Economics*, 79, 72–90. doi: 10.1016/j.jue.2013.03.007
- Craig, L. A., Palmquist, R. B., & Weiss, T. (1998). Transportation Improvements and Land Values in the Antebellum United States: A Hedonic Approach. *Journal of Real Estate Finance and Economics*, 16(2), 173–189. doi: 10.1023/A:1007755717900
- Dai, R., Mookherjee, D., Munshi, K., & Zhang, X. (2020). *The community origins of private enterprise in china* (Discussion Paper No. 320). Boston University, Institute for Economic Development.
- Dai, R., Mookherjee, D., Munshi, K., & Zhang, X. (2023). *Entrepreneurship in china’s structural transitions: Network expansion and overhang* (Working Paper No. 31477). National Bureau of Economic Research. doi: 10.3386/w31477
- Desmet, K., Nagy, D. K., & Rossi-Hansberg, E. (2018). The geography of development. *Journal of Political Economy*, 126(3), 903–983. doi: 10.1086/697084
- Dix-Carneiro, R. (2014). Trade liberalization and labor market dynamics. *Econometrica*, 82(3), 825–885. doi: 10.3982/ECTA11214
- Donaldson, D. (2018). Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. *American Economic Review*, 108(4-5), 899–934. doi: 10.1257/aer.20101199
- Donaldson, D., & Hornbeck, R. (2016, 02). Railroads and american economic growth: A “market access” approach *. *The Quarterly Journal of Economics*, 131(2), 799–858.
- Donovan, K., & Schoellman, T. (2023). The role of labor market frictions in structural transformation. *Oxford Development Studies*, 51(4), 362–374. doi: 10.1080/13600818.2023.2276702
- Eaton, J., & Kortum, S. (2002). Technology, Geography, and Trade. *Econometrica*, 70(5), 1741–1779. doi: 10.1111/1468-0262.00352
- Eckert, F., Gvirtz, A., Liang, J., & Peters, M. (2020, February). *A method to construct geographical crosswalks with an application to us counties since 1790* (Working Paper No. 26770). National Bureau of Economic Research.
- Eckert, F., & Kleineberg, T. (2024). *The geography of opportunity: Education, work, and intergenerational mobility across us counties* (Institute Working Paper No. 47). Federal Reserve Bank of Minneapolis, Opportunity & Inclusive Growth Institute. (Revise and Resubmit at the Journal of Political Economy) doi: 10.21034/iwp.47

- Eckert, F., & Peters, M. (2022). *Spatial structural change* (NBER Working Paper No. 30489). National Bureau of Economic Research. doi: 10.3386/w30489
- Enke, B. (2019). Kinship, Cooperation, and the Evolution of Moral Systems. *Quarterly Journal of Economics*, 134(2), 953–1019. doi: 10.1093/qje/qjz001
- Espín-Sánchez, J.-A., Ferrie, J. P., & Vickers, C. (2023, August). *Women and the econometrics of family trees* (Working Paper No. 31598). National Bureau of Economic Research.
- Fan, J., & Zou, B. (2021). *The dual local markets: Family, jobs, and the spatial distribution of skills*. (Working paper)
- Feigenbaum, J. J. (2018). Multiple measures of historical intergenerational mobility: Iowa 1915 to 1940. *Economic Journal*, 128(612), F446–F481. doi: 10.1111/ecoj.12513
- Fishback, P. (2017). How Successful Was the New Deal? The Microeconomic Impact of New Deal Spending and Lending Policies in the 1930s. *Journal of Economic Literature*, 55(4), 1435–1485.
- Fogel, R. W. (1964). *Railroads and American Economic Growth: Essays in Econometric History*. Baltimore: Johns Hopkins University Press.
- Fogli, A., & Guerrieri, V. (2019, August). *The end of the american dream? inequality and segregation in us cities* (Working Paper No. 26143). National Bureau of Economic Research.
- Fryer, J., Roland G., & Levitt, S. D. (2004, 08). The causes and consequences of distinctively black names. *The Quarterly Journal of Economics*, 119(3), 767–805.
- Ghosh, A., Hwang, S. I. M., & Squires, M. (2023). Economic consequences of kinship: Evidence from U.S. bans on cousin marriage. *Quarterly Journal of Economics*, 138(4), 2559–2606. doi: 10.1093/qje/qjad018
- Glitz, A. (2017). Coworker networks in the labour market. *Labour Economics*, 44, 218–230. doi: 10.1016/j.labeco.2016.12.006
- Gouriéroux, C., Monfort, A., & Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics*, 8(S1), S85–S118. doi: 10.1002/jae.3950080507
- Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6), 1360–1380. doi: 10.1086/225469
- Helgertz, J., Price, J., Wellington, J., Thompson, K. J., Ruggles, S., & Fitch, C. A. (2022). A New Strategy for Linking U.S. Historical Censuses: A Case Study for the IPUMS Multigenerational Longitudinal Panel. *Historical Methods*, 55(1), 12–29.
- Helgertz, J., Ruggles, S., Warren, J. R., Fitch, C. A., Hacker, J. D., Nelson, M. A., ... Sobek, M. (2023). *IPUMS Multigenerational Longitudinal Panel: Version 1.1 [dataset]* (Tech. Rep.). IPUMS.
- Henrich, J. (2020). *The WEIRD people in the world: How the west became psychologically peculiar and particularly prosperous*. New York: Farrar, Straus and Giroux.

- Herrendorf, B., & Schoellman, T. (2018). Wages, human capital, and barriers to structural transformation. *American Economic Journal: Macroeconomics*, 10(2), 1–23. doi: 10.1257/mac.20160236
- Hornbeck, R. (2012). The enduring impact of the american dust bowl: Short- and long-run adjustments to environmental catastrophe. *American Economic Review*, 102(4), 1477–1507.
- Hornbeck, R. (2023). Dust bowl migrants: Environmental refugees and economic adaptation. *The Journal of Economic History*, 83(3), 645–675. doi: 10.1017/S0022050723000244
- Hornbeck, R., & Rotemberg, M. (2024). Growth off the rails: Aggregate productivity growth in distorted economies. *Journal of Political Economy*, 132(11), 3547–3602. doi: 10.1086/730548
- Ilyés, V., Boza, I., Lőrincz, L., & Eriksson, R. H. (2023). How to enter high-opportunity places? the role of social contacts for residential mobility. *Journal of Economic Geography*, 23(2), 371–395. doi: 10.1093/jeg/lbac019
- Ioannides, Y. M., & Loury, L. D. (2004). Job information networks, neighborhood effects, and inequality. *Journal of Economic Literature*, 42(4), 1056–1093. doi: 10.1257/0022051043004595
- Jäger, S., & Heining, J. (2022). *How substitutable are workers? evidence from worker deaths* (Working Paper No. 30629). National Bureau of Economic Research. doi: 10.3386/w30629
- Jaravel, X., Petkova, N., & Bell, A. (2018). Team-specific capital and innovation. *American Economic Review*, 108(4–5), 1034–1073. doi: 10.1257/aer.20151184
- Jones, B., & Olken, B. (2005). Do leaders matter? national leadership and growth since world war ii. *The Quarterly Journal of Economics*, 120(3), 835–864.
- Kaplanis, J., Gordon, A., Shor, T., Weissbrod, O., Geiger, D., et al. (2018). Quantitative Analysis of Population-Scale Family Trees with Millions of Relatives. *Science*, 360(6385), 171–175.
- Kennan, J., & Walker, J. R. (2011). The effect of expected income on individual migration decisions. *Econometrica*, 79(1), 211–251. doi: 10.3982/ECTA4657
- Kim, J. S., Patacchini, E., Picard, P. M., & Zenou, Y. (2023). Spatial interactions. *Quantitative Economics*, 14(4), 1295–1335. doi: 10.3982/QE1720
- Koenen, M., & Johnston, D. (2024). *Social ties and residential choice: Micro evidence and equilibrium implications*. (Working paper)
- Koylu, C., Guo, D., Huang, Y., Kasakoff, A., & Grieve, J. (2021). Connecting family trees to construct a population-scale and longitudinal geo-social network for the u.s. *International Journal of Geographical Information Science*, 35(12), 2380–2423. doi: 10.1080/13658816.2020.1821885

- Kramarz, F., & Nordström Skans, O. (2014). When strong ties are strong: Networks and youth labour market entry. *Review of Economic Studies*, 81(3), 1164–1200. doi: 10.1093/restud/rdt049
- Kuznets, S. (1973, June). Modern economic growth: Findings and reflections. *American Economic Review*, 63(3), 247–258.
- Lee, D., & Wolpin, K. I. (2006). Intersectoral Labor Mobility and the Growth of the Service Sector. *Econometrica*, 74(1), 1–46.
- Lin, X., & Mo, S. (2024). *Does mobility beget mobility? coworker networks and the sectoral reallocation of labor* (Working Paper).
- Lindahl, M., Palme, M., Sandgren Massih, S., & Sjögren, A. (2015). Long-term intergenerational persistence of human capital: An empirical analysis of four generations. *Journal of Human Resources*, 50(1), 1–33. doi: 10.3368/jhr.50.1.1
- Liu, J.-Y. (2026). Natural disaster and intergenerational occupational choice in agriculture: Evidence from the us dust bowl. *American Journal of Agricultural Economics*. (Accepted / Early View) doi: 10.1002/ajae.70081
- Lo Bello, S., & Morchio, I. (2022). Like father, like son: Occupational choice, intergenerational persistence and misallocation. *Quantitative Economics*, 13(2), 629–679. doi: 10.3982/QE1375
- Long, J., & Ferrie, J. (2013). Intergenerational occupational mobility in great britain and the united states since 1850. *American Economic Review*, 103(3), 1109–1137. doi: 10.1257/aer.103.3.1109
- McKenzie, D., & Rapoport, H. (2007). Network effects and the dynamics of migration and inequality: Theory and evidence from mexico. *Journal of Development Economics*, 84(1), 1–24. doi: 10.1016/j.jdeveco.2006.11.003
- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27, 415–444. doi: 10.1146/annurev.soc.27.1.415
- Miller, G. (2006). *Vital statistics deaths: Historical 1900–1936*. National Bureau of Economic Research. Retrieved from <https://www.nber.org/research/data/vital-statistics-deaths-historical-1900-1936> (Digitized from the annual U.S. Bureau of the Census Mortality Statistics volumes)
- Montgomery, J. D. (1991). Social networks and labor-market outcomes: Toward an economic analysis. *American Economic Review*, 81(5), 1407–1418.
- Morten, M. (2019). Temporary migration and endogenous risk sharing in village india. *Journal of Political Economy*, 127(1), 1–46. doi: 10.1086/700763
- Munshi, K. (2003, 05). Networks in the modern economy: Mexican migrants in the u. s. labor market*. *The Quarterly Journal of Economics*, 118(2), 549–599.
- Munshi, K. (2011). Strength in Numbers: Networks as a Solution to Occupational Traps.

- Review of Economic Studies*, 78(3), 1069–1101. doi: 10.1093/restud/rdq029
- Munshi, K. (2020). Social Networks and Migration. *Annual Review of Economics*, 12, 503–524.
- Munshi, K., & Rosenzweig, M. (2016). Networks and misallocation: Insurance, migration, and the rural-urban wage gap. *American Economic Review*, 106(1), 46–98. doi: 10.1257/aer.20131365
- Ngai, L. R., & Pissarides, C. A. (2007, March). Structural change in a multisector model of growth. *American Economic Review*, 97(1), 429–443.
- Nicoletti, C., Salvanes, K. G., & Tominey, E. (2018). The Family Peer Effect on Mothers' Labor Supply. *American Economic Journal: Applied Economics*, 10(3), 206–234.
- Oettl, A. (2012). Reconceptualizing stars: Scientist helpfulness and peer performance. *Management Science*, 58(6), 1122–1140. doi: 10.1287/mnsc.1110.1470
- Patel, K., & Vella, F. (2013). Immigrant networks and their implications for occupational choice and wages. *Review of Economics and Statistics*, 95(4), 1249–1277. doi: 10.1162/REST_a_00327
- Porzio, T., Rossi, F., & Santangelo, G. (2022). The human side of structural transformation. *American Economic Review*, 112(8), 2774–2814. doi: 10.1257/aer.20201157
- Raz, I. T. (2025). Soil heterogeneity, social learning, and the formation of close-knit communities. *Journal of Political Economy*, 133(8), 2643–2691. doi: 10.1086/735506
- Redding, S. J., & Rossi-Hansberg, E. (2017). Quantitative spatial economics. *Annual Review of Economics*, 9, 21–58. doi: 10.1146/annurev-economics-063016-103713
- Rees, A. (1966). Information networks in labor markets. *American Economic Review (Papers and Proceedings)*, 56(1/2), 559–566.
- Roberts, S. G. B., & Dunbar, R. I. M. (2011). The costs of family and friends: An 18-month longitudinal study of relationship maintenance and decay. *Evolution and Human Behavior*, 32(3), 186–197. doi: 10.1016/j.evolhumbehav.2010.08.005
- Rodríguez-Clare, A., Ulate, M., & Vásquez, J. P. (2026). Trade with nominal rigidities: Understanding the unemployment and welfare effects of the china shock. *Journal of Political Economy*, 134(2), 626–664. doi: 10.1086/738344
- Ruggles, S., Fitch, C. A., & Roberts, E. (2018). Historical census record linkage. *Annual Review of Sociology*, 44, 19–37. doi: 10.1146/annurev-soc-073117-041447
- Ruggles, S., Nelson, M. A., Sobek, M., Fitch, C. A., Goeken, R., Hacker, J. D., ... Warren, J. R. (2024). *Ipums ancestry full count data: Version 4.0 [dataset]*. IPUMS. doi: 10.18128/D014.V4.0
- Sahai, H., & Bailey, M. (2023). *Social Networks and Internal Migration: Evidence from Facebook in India* (Tech. Rep.). Working Paper.
- Schroeder, J., Riper, D. V., Manson, S., Knowles, K., Kugler, T., Roberts, F., & Ruggles, S. (2025). *Ipums national historical geographic information system: Version 20.0 [dataset]*.

- IPUMS. doi: doi.org/10.18128/D050.V20.0
- Schulz, J. F. (2022). Kin networks and institutional development. *Economic Journal*, 132(647), 2578–2613. doi: 10.1093/ej/ueac027
- Schulz, J. F., Bahrami-Rad, D., Beauchamp, J. P., & Henrich, J. (2019). The church, intensive kinship, and global psychological variation. *Science*, 366(6466), eaau5141. doi: 10.1126/science.aau5141
- Silva, J. M. C. S., & Tenreyro, S. (2006, 11). The log of gravity. *The Review of Economics and Statistics*, 88(4), 641–658.
- Stuart, B. A., & Taylor, E. J. (2021, July). Migration networks and location decisions: Evidence from us mass migration. *American Economic Journal: Applied Economics*, 13(3), 134–75.
- Takeda, K. (2022). *The geography of structural transformation: Effects on inequality and mobility* (CEP Discussion Paper No. 1893). Centre for Economic Performance, London School of Economics.
- Taylor, G. R. (1951). *The Transportation Revolution, 1815–1860* (Vol. 4). New York: Rinehart.
- Tolbert, C. M., & Sizer, M. (1996, September). *U.S. Commuting Zones and Labor Market Areas: A 1990 Update* (Staff Paper No. 9614). USDA Economic Research Service.
- U.S. Bureau of Labor Statistics. (1949). *Occupational outlook handbook: Employment information on major occupations for use in guidance* (Bulletin No. 940). Washington, DC: U.S. Department of Labor, Bureau of Labor Statistics.
- Ward, Z. (2023, December). Intergenerational mobility in american history: Accounting for race and measurement error. *American Economic Review*, 113(12), 3213–48.
- Wing, C., Freedman, S. M., & Hollingsworth, A. (2024, January). *Stacked difference-in-differences* (Working Paper No. 32054). National Bureau of Economic Research. doi: 10.3386/w32054
- Yang, D. (2024). *A general equilibrium investigation of the american dust bowl* (SocArXiv No. jxz2b). Center for Open Science. (Revise and Resubmit at the Journal of Political Economy Microeconomics) doi: 10.31235/osf.io/jxz2b
- Zabek, M. (2024, April). Local ties in spatial equilibrium. *American Economic Journal: Macroeconomics*, 16(2), 287–317. doi: 10.1257/mac.20210326
- Zerecero, M. (2021). *The birthplace premium*. (Working paper)
- Zhang, L., Xie, L., & Zheng, X. (2023). Across a few prohibitive miles: The impact of the anti-poverty relocation program in china. *Journal of Development Economics*, 160, 102945. doi: 10.1016/j.jdeveco.2022.102945

APPENDIX

A1	Family ties construction	59
A1.1	Detailed procedure	59
A1.2	Internal validation	70
A1.3	External benchmarks for the family network measures	76
A1.4	Summary statistics of families	79
A1.5	External benchmark for inferred deaths	81
A1.6	Summary statistics of death variables	82
A2	Empirics	83
A2.1	The evolution of family networks across sectors and space	83
A2.2	Kinship and place-based social connections	84
A2.3	Family connections and sectoral choice	88
A2.4	Families and migration choice	103
A3	Theory	107
A3.1	Stylized model	107
A3.2	Model extensions	113
A4	Quantification	118
A4.1	Data	119
A4.2	Construction of migration distance	122
A4.3	Estimation of migration costs	123
A4.4	Sector deaths and continuation-value adjustment	126
A4.5	Indirect Inference	129
A4.6	Model inversion	136
A4.7	Representative-kin approximation	145
A5	Quantitative analysis	150
A5.1	Solution algorithm	150
A5.2	Additional aggregate effects of endogenous kin adjustment	154
A5.3	The initial horizontal-kin distribution	154
A5.4	Endogenous propagation of horizontal kin	156
A5.5	Relocation voucher	163

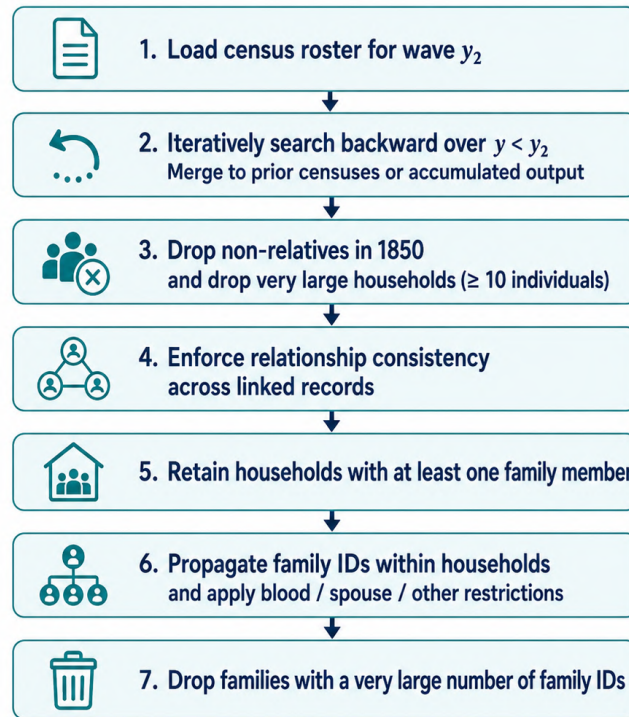
A1 Family ties construction

A1.1 Detailed procedure

This appendix describes in detail how I construct multigenerational family networks for the United States using decennial census waves from 1850 to 1940. The objective is to assign each individual observed in a given census wave a set of extended-family identifiers that can cross household and region boundaries, while maintaining internal consistency with observed household rosters and relationship codes. The key challenge is that households can contain multiple lineages (e.g., married children co-residing with parents, widowed parents joining adult children, siblings living together) and a given individual may simultaneously belong to multiple family groups (e.g., through marriage). The procedure therefore iteratively links individuals across census waves, anchors households to previously observed family identifiers and propagates those identifiers within households only when the implied relationship structure is consistent. The algorithm proceeds sequentially by census wave. For each target wave from 1860–1940, it searches backward across earlier waves using crosswalk files that map individuals to their prior-wave individual identifiers. Whenever a match is found, the algorithm retrieves the individual’s last observed household identifier and relationship code in the earlier wave. These inherited identifiers become anchors for constructing extended family memberships in the current census wave. Because a single erroneous link can mechanically contaminate an entire household if family identifiers are propagated without checks, the procedure is deliberately conservative. It implements (i) early-wave restrictions in 1850, (ii) a relationship-consistency check that compares inferred kinship structure across the old and new household rosters and (iii) a set of propagation rules that restrict how family identifiers can spread through the household when the link connecting households is via a spouse, via a descendant (“linking from below”), or via collateral relatives (e.g., cousins or nephews). Figure [A1.1.1](#) summarizes the overall construction pipeline described below.

Data inputs. The construction combines three inputs. The first is the full-count census in each wave taken from IPUMS. For every person, the census identifies the historical identifier (`histid`), household identifier (`serial`), the individual’s relationship to the household head (`relate`) and, when available, links to co-resident parents (`momloc` for mothers and `poploc` for fathers). The second input is the set of Census Tree Project crosswalks, which link individuals across census waves ([Buckles et al., 2025](#)). The third input is a relationship-pair crosswalk that I construct, as described below. This crosswalk maps the relationships of two household members to the head into their relationship to each other. For example, if one person is coded as a child of the head and another as the parent of the head, the crosswalk classifies the first person as the grandchild of the second. When a pair of relationship codes is not uniquely

Figure A1.1.1: Steps for constructing multigenerational family ties



Notes: This figure describes the construction of the family ties. Individuals are linked backward across censuses, inconsistent or implausible relationships are removed and family identifiers are propagated to construct multigenerational networks.

informative, I retain an ambiguity flag.

Iterative backward linkage and anchoring. For each target census wave, I search backward through all earlier waves for linked relatives. The set of possible source waves expands over time: for the 1860 census I search only 1850, while for the 1940 census I search 1930, 1920, 1910, 1900, 1880, 1870, 1860 and 1850. I start from the full roster in the target wave and create fields that record the first successful historical anchor: the linked person in an earlier census, the year of that link, the household in which the person was previously observed and the person's relationship to the head in that earlier household.

The backward search proceeds year by year. When a target-wave individual is linked to an earlier, more recent census through the crosswalk, I retrieve the family information already attached to that earlier observation. For links to 1850, this information comes directly from the raw 1850 household structure. For links to later source years, it comes from the family-ties file that has already been constructed for that year. Thus, family labels accumulate recursively: a person linked to 1910 can inherit family groups that were themselves built from links to 1900,

1880 and earlier waves.

Each individual is assigned at most one primary historical anchor within a target wave. Once a valid backward match is found, that person is removed from later searches over older waves. This rule avoids conflicts when an individual has multiple possible cross-census links. Additional family connections are still added through within-household relationships and marriage links, but the historical anchor used to bring forward inherited family labels is unique.

Early-wave restrictions (1850). The 1850 census provides limited relationship detail. In particular, several categories of co-residents are not consistently classifiable as blood relatives in a way that is stable under forward propagation. To avoid introducing spurious kin links at the baseline of the panel, I drop from the set of anchored individuals those whose last observed household is in 1850 and who are coded as non-relatives (`RELATE` codes ≥ 1061) or as “other relatives” (`RELATE` code 1001). This restriction is intentionally conservative: it reduces the risk that ambiguous household ties in 1850 generate false extended family connections in later waves.

Relationship consistency check. A central concern in this linking procedure is that a single incorrect person-level link can cause an entire household to inherit incorrect family identifiers. The algorithm therefore implements a relationship-consistency check designed to detect mismatches between the implied kinship structure of the old household and the observed structure of the new household.

The check proceeds as follows. Within each new household in a given census wave, consider the subset of individuals who were successfully linked to some prior household. Among those linked individuals, the algorithm selects a reference person based on a deterministic “seniority” ordering of relationship codes that prioritizes older generations and household heads or spouses. Table A1.1.2 presents the ordering. Intuitively, the most senior linked person is the most informative anchor because their relationships to other household members are typically measured with less ambiguity.

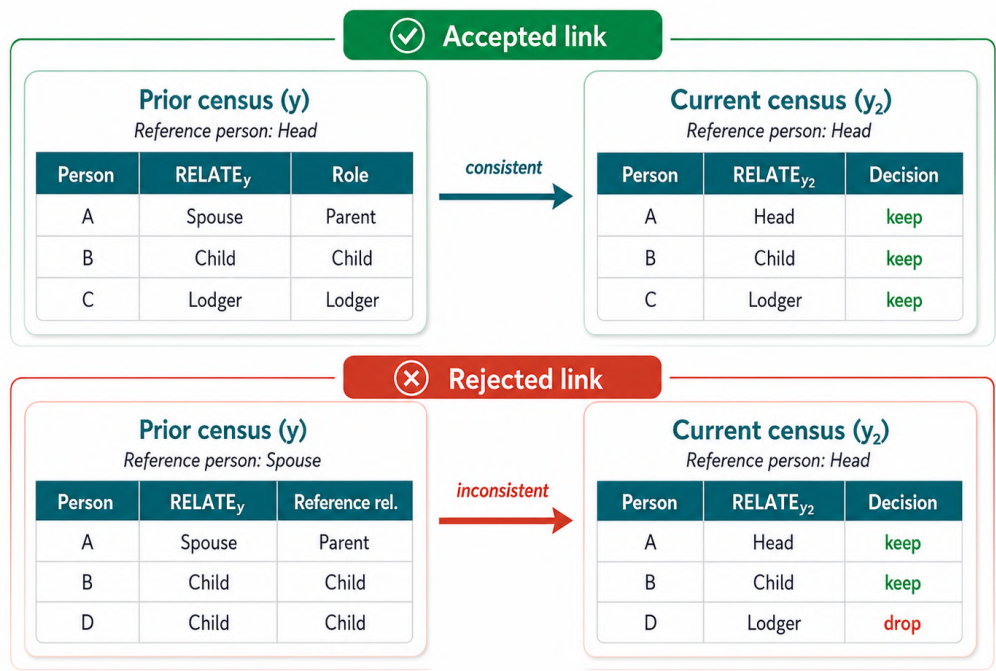
Once the reference person is determined, the algorithm constructs two mappings for every person in the household. The first mapping uses the linked individuals’ prior-wave relationship codes to infer each person’s relationship to the reference person in the old household. The second mapping uses the current-wave relationship codes to infer each person’s relationship to the reference person in the new household. Both mappings rely on the constructed relationship-pair crosswalk, which converts a pair of relationships to the household head into an inferred relationship between the two individuals, together with an ambiguity flag when the relationship is not uniquely determined (e.g., “other relatives”).

A merge is retained if the inferred relationship of each individual to the reference person agrees across the old and new mappings *and* the individuals originate from the same prior household,

or if the individual is the linking person itself, allowing household composition to change. The procedure does not reject cases in which either side of the relationship mapping is missing, since missing mappings frequently correspond to boarders or ambiguous codes for which the crosswalk cannot uniquely determine kinship, or to cases in which the relationship cannot be measured because the individuals were present in different households earlier.

Figure A1.1.2 illustrates how the relationship-consistency check works in a stylized example.

Figure A1.1.2: Relationship-consistency check: accepted vs. rejected links



Notes: The algorithm enforces stability of inferred kin roles across linked census waves. Links are retained when role changes are uninformative (e.g., lodgers) but rejected when a previously identified family member appears with an incompatible role.

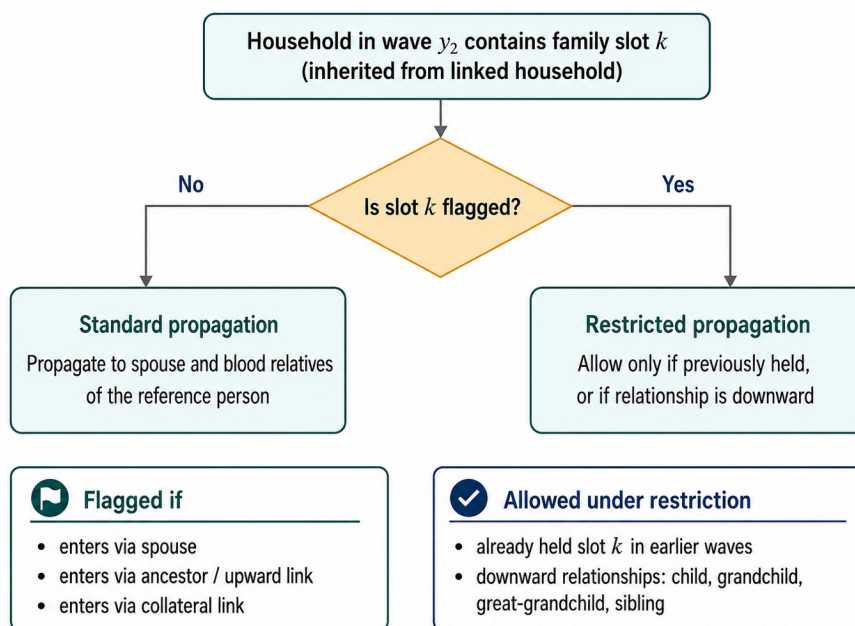
The logic is that the procedure is stringent when it matters: when an individual has a strong implied kin relationship in the old household (e.g., child, parent, sibling) but appears inconsistent with the structure of the new household. By contrast, when relationship inference is missing on either side—often because the individual is a lodger or has an ambiguous code—the procedure does not force rejection, since such records are not used as anchors for family propagation.

Propagation of family identifiers within households. After the consistency check, the algorithm retains only households in which at least one member carries an inherited family

identifier. This household-level propagation allows family trees to expand over time while ensuring that every family label assigned in a new wave is anchored to at least one credible backward link. The reference anchor is chosen first by the presence of an inherited family identifier and then by seniority. Within the remaining households, the procedure enumerates the number of distinct inherited prior-household identifiers present. In practice, this number can exceed one: a household can contain multiple linked lineages.

Propagation is then carried out family-slot by family-slot. For each slot k , the algorithm defines a reference person within the new household and uses the relationship-pair crosswalk to infer whether each member of the household is a blood relative or spouse of that reference person. Individuals inherit slot k if they are classified as a blood relative or spouse, or if they already carried that family identifier from prior waves. Individuals who are neither blood relatives nor spouses and who did not previously carry that identifier do not inherit it. This ensures that family identifiers do not mechanically spill over to boarders, servants, or unrelated co-residents.

Figure A1.1.3: Within household propagation



Notes: This figure shows how the family ties are linked within households. Family identifiers propagate to spouses and blood relatives; flagged links are retained only when previously observed or permitted by the relationship direction.

Cleaning propagation. Even when a household contains a valid historical anchor, not all anchors are equally informative for forward propagation across generations. The code tracks three special cases in which the connecting relationship is likely to create spurious upward or

cross-lineage assignments if used without restriction.

First, I flag cases where the family slot enters the household through a spousal tie. Such links are informative for marriage connections but, without additional information, can blur lineage boundaries if used to infer broader blood relationships of co-residents, for example parents-in-law or cousins of the household head. Second, I flag cases where the connection is through descendants (grandchildren or great-grandchildren). These links can generate ambiguous upward propagation because descendants may carry a family identifier inherited through a parent who is not connected to the lineage of the grandparent used as the current anchor. Third, I flag collateral links such as cousins, nephews or nieces and grand-nieces or grand-nephews, where extending the identifier further can similarly blur lineage boundaries.

If a family slot is flagged by any of these indicators, propagation in the next iteration is restricted: individuals who would inherit the slot solely through an upward or collateral mapping are prevented from doing so unless they are themselves in downward relationships (child, grandchild, great-grandchild, or sibling) or already carried the identifier historically.

These restrictions are designed to preserve lineage membership rather than co-residence. Spousal, descendant and collateral links can identify valid kin connections, but they are also the cases in which a small linkage error can generate implausible assignments across several generations. Restricting their subsequent propagation therefore trades some coverage for greater precision. I additionally require cross-census links to be supported by at least two of the linking specifications incorporated in the Census Tree Project. The resulting files record each individual's inherited family identifiers and position within those lineages, which form the basis for the sectoral, spatial and relationship-specific kin measures used in the analysis.

Table [A1.1.1](#) provides a summary of the number of linkages after the cleaning.

A1.1.1 Inferring pairwise relationships and tree positions

The construction described above assigns each individual to one or more lineages, but it does not, on its own, record how co-residents are related to one another or where each sits in a family tree. Two objects supply this information. The first is a relationship-pair crosswalk that infers the relationship between *any* two household members from their relationship to the head. The second is a compact tree-position encoding that records each individual's genealogical path from the household head and carries that path forward across census waves. The crosswalk is what makes the relationship between non-head co-residents observable at all; the path encoding is what turns those pairwise relationships into a coherent, multigenerational tree.

The relationship-pair crosswalk. The IPUMS `related` variable codes each person's relationship to the household head only, so the relationship between two non-head members is not given directly. I therefore pre-compute, once, a crosswalk over all ordered pairs of `related`

Table A1.1.1: Linkage and cleaning summary by census wave

Wave	Raw roster	Drop non-rel. in 1850 & large HH	Drop incons. rel. check	Link within HH	Within HH rel. check	Final ties
1860	27,451,053	24,922,901	24,623,640	13,837,017	12,458,274	12,458,274
1870	38,515,225	35,470,487	35,161,290	15,055,802	13,175,659	13,175,659
1880	50,140,482	46,686,457	46,293,189	18,991,034	17,464,960	17,464,295
1900	75,824,712	70,986,557	70,654,648	24,863,913	21,971,465	21,961,359
1910	92,043,618	85,822,285	85,197,264	28,129,945	24,802,154	24,728,487
1920	105,730,985	100,614,636	99,883,063	31,319,830	28,077,928	27,815,220
1930	122,777,512	117,731,763	116,879,915	36,204,604	32,287,648	31,411,504
1940	131,902,398	127,927,495	127,070,751	37,344,793	33,562,530	32,213,482

Notes: “Raw roster” corresponds to the initial roster size in each census wave. “Drop non-rel. in 1850 & large HH” applies the early-wave filtering rule for 1850 (non-relatives/other relatives) and large-household exclusions. “Drop incons. rel. check” applies the crosswalk-based relationship consistency filter. “Link within HH” reports the number of individuals remaining in households with at least one family member (post household-presence filter). “Within HH rel. check” corresponds to the post across-time linkage count and “Final ties” is the final sample after all cleaning steps.

codes. For each pair (rel_i, rel_r) —person i ’s code to the head and reference person r ’s code to the head—the crosswalk returns i ’s relationship to r together with an ambiguity flag. Two distributions are exact by construction: when the reference is the head, i ’s relationship to r is simply i ’s own **related** code; and when i is the head, i ’s relationship to r is the genealogical reciprocal of r ’s code (the head is the parent of the head’s child, the child of the head’s parent and so on). For all remaining pairs, the crosswalk applies genealogical logic organized by the reference person’s role. If the reference is a child of the head, then another child of the head is a sibling of the reference, the head’s parent is a grandparent of the reference, the head’s sibling is an aunt or uncle of the reference and a grandchild of the head is a niece or nephew of the reference; analogous rules cover references who are spouses, parents, siblings, grandparents, aunts or uncles, nieces or nephews, cousins and the in-law counterparts. Wherever the **related** codes allow, the crosswalk distinguishes blood, step/half/adopted and in-law variants.

Ambiguity. Because it uses only relationship-to-head codes, the crosswalk cannot resolve every pair uniquely. When the head has several children, a child-in-law cannot be matched to the specific child he or she married; a second person coded as “parent” of the head is most plausibly the head’s other parent (hence the spouse of the first), but this is not certain; and a step or adopted code on either side turns an inferred “sibling” into a half- or step-sibling. Such cases receive an ambiguity flag and pairs that cannot be resolved at all default to the “other

Table A1.1.2: Seniority ordering for linking

Seniority	RELATE code	IPUMS label
Panel A. Core household roles		
1	1011	Grandparent
2	0501	Parent
3	1021	Aunt or uncle
4	0101	Head/householder
5	0201	Spouse
6	0701	Sibling
7	0301	Child
8	1041	Cousin
9	1031	Nephew, niece
10	0901	Grandchild
Panel B. Step relations		
11	1012	Step grandparent
12	0502	Step parent
13	0702	Step/half/adopted sibling
14	0303	Stepchild
15	0903	Step grandchild
16	1033	Step/adopted nephew/niece
Panel C. In-laws (affinal ties)		
17	1013	Grandparent-in-law
18	1022	Aunt-/uncle-in-law
19	0601	Parent-in-law
20	0602	Step parent-in-law
21	0801	Sibling-in-law
22	0802	Step/half sibling-in-law
23	0401	Child-in-law
24	0402	Step child-in-law
25	0904	Grandchild-in-law
26	1034	Grand niece/nephew

Notes: This table documents the deterministic mapping used to construct the variable seniority. Seniority is assigned from the IPUMS RELATE code (here shown with leading zeros) and is used to prioritize individuals for linking and for tie-propagation ordering. Lower values denote higher priority.

relative” category. Throughout the construction, ambiguous and non-blood pairs are treated conservatively: they may be retained as observed ties, but they are not used to extend blood lineages upward under the propagation rules above, which prevents a single uncertain inference from cascading through the tree. The same crosswalk is used for the relationship-consistency check: it allows me to compute and then compare each member’s relationship to the senior reference person in both the origin and destination household.

Tree positions as genealogical paths. To make the full kin structure explicit, I assign every individual a position in the family tree, written as a path from the household head. A position is a sequence of elementary steps: *p* moves up to a parent, *c* moves down to a child and *s* moves across to a spouse, with the head as the root (“0”). Direct relatives use a single step (spouse *s*, child *c*, parent *p*); relatives reached only by traversing an intermediate node that is not itself observed carry a “?” for that node. Table A1.1.3 lists the encoding: a sibling is “*p?.c*” (up to an unobserved shared parent, then down to a child), a grandparent “*p?.p*”, an aunt or uncle “*p?.p?.c*”, a niece or nephew “*p?.c?.c*”, a cousin “*p?.p?.c?.c*” and a great-grandchild “*c?.c?.c*”; in-laws append a spousal step, so a child-in-law is “*c?.s*” and a parent-in-law “*s?.p*”. A terminal index equal to the person number within the household distinguishes, for example, one sibling or child from another.

Carrying positions forward. Positions are seeded in 1850, where each person is placed relative to the head of their 1850 household. When a household is linked to a later wave, members who are already in the tree keep their existing position and newly observed co-residents are assigned a position by appending the step implied by the crosswalk to the position of the linked reference person. A single chain of links can thus build up multi-step paths—a grandchild first observed in a later wave inherits the reference’s position as a prefix and adds the steps down to its own generation—so paths lengthen and the tree deepens as waves accumulate from 1850 to 1940.

Pairwise relationships and final cleaning. The tree positions and generation indices allow me to recover the relationship between any two individuals who share a lineage, rather than only their relationship to the household head. I classify pairs using their relative position in the tree and the number of generations separating them. This identifies spouses, parents and children, siblings, grandparents and grandchildren, aunts and uncles, nieces and nephews and cousins. Because a spousal step and a blood step both count as a single link in the tree, the resulting categories include the corresponding in-law relationships; for example, parents-in-law are grouped with parents and siblings-in-law with siblings.

I apply three additional cleaning rules to these pairwise relationships. First, the same pair can be connected through more than one lineage. I retain each unordered pair only once, using the lineage that implies the shortest genealogical path between them. This prevents overlapping lineage identifiers from mechanically inflating kin counts. Second, more distant relationships are retained only when both sides of the connection can be traced through direct lineage links rather than through previously flagged spousal, descendant or collateral propagation. These are the distributions in which small linkage errors are most likely to accumulate across generations. Third, I remove distant relationships that imply implausible age differences given the inferred kinship category. Pairs with missing generation information or unresolved tree positions are

left unclassified and do not enter the relationship-specific measures. Appendix [A1.2](#) validates these restrictions.

Table A1.1.3: Tree-position path encoding

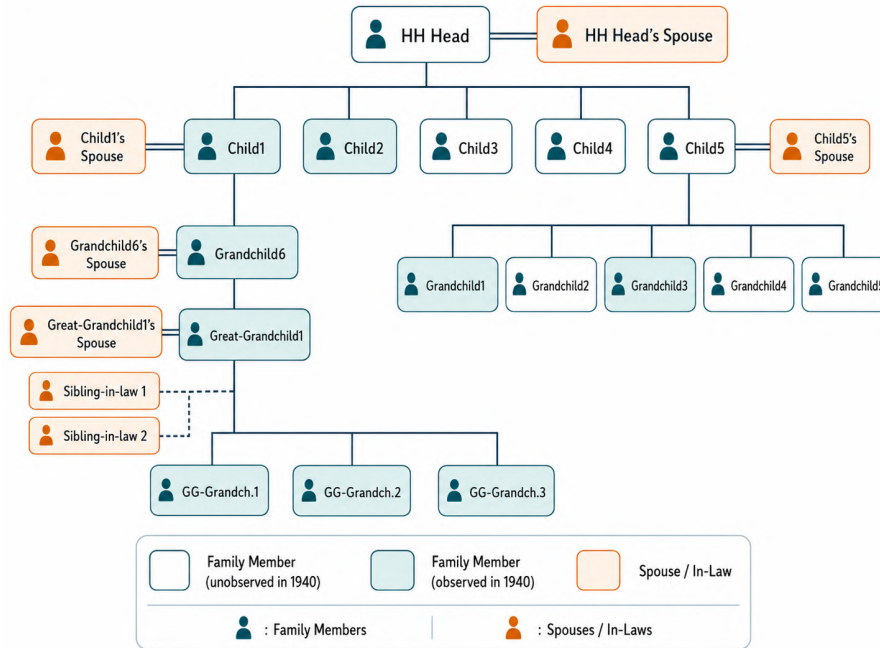
Relationship to reference	Tree path
Head (root)	0
Spouse	s
Child	c
Parent	p
Sibling	p?.c
Grandchild	c?.c
Grandparent	p?.p
Aunt or uncle	p?.p?.c
Niece or nephew	p?.c?.c
Cousin	p?.p?.c?.c
Great-grandchild	c?.c?.c
Grand-niece/nephew	p?.c?.c?.c
Child-in-law	c?.s
Parent-in-law	s?.p
Sibling-in-law	p?.c?.s

Notes: Each individual’s tree position is a path from the household head built from three elementary steps: **p** (up to a parent), **c** (down to a child) and **s** (across to a spouse), with the head as the root (0). A “?” marks an intermediate node that is inferred from the relationship-pair crosswalk rather than directly observed in the household roster; a sibling (**p?.c**) is thus reached by going up to a shared parent and back down to a child. A terminal index equal to the person number distinguishes, e.g., one sibling from another. Step, half, adopted and in-law variants share the same path skeleton as their blood counterparts. The mapping reproduces the assignment used to build the `tree_step` variable in the construction code.

A1.1.2 Reweighting linked individuals

This appendix describes how I reweight the linked family data to reduce observable differences associated with selective linkage. Because family ties can only be constructed for individuals who are successfully linked across censuses, the resulting sample is not a random draw from the population. Linkage success varies systematically with demographic characteristics, location and socioeconomic status. Ignoring this selection would bias any estimates that use family structure or network composition. To address this issue, I reweight the linked sample using an inverse propensity-score approach that balances it toward the full census along observed

Figure A1.1.4: Family tree example



Notes: This figure is an illustrative example of a reconstructed family tree descending from an ancestor head. Children and spouses are organized hierarchically. Teal boxes denote individuals present in the 1940 census. Orange boxes denote individuals who are not members of the focal household head's blood lineage; they may nevertheless belong to other reconstructed family lineages.

characteristics.

The procedure begins by stacking, for each census year, the linked sample (individuals for whom family ties are observed) together with the full underlying census population in that year. Individuals are classified according to whether they are observed in the linked sample or not. I then estimate, separately by census year, a reduced-form model for the probability that an individual is successfully linked. The specification conditions on a rich set of pre-determined individual characteristics that are observed for both linked and unlinked individuals. These include demographic attributes (age group, sex, race, marital status), birthplace and state of residence, urban versus rural status, sector and skill group, industry sector, farming status, literacy (when available) and indicators for internal migration and immigration. All of these variables are measured contemporaneously in the census and are plausibly related both to linkage success and to family structure, but not affected by the subsequent outcomes studied in the paper.

Using this model, I compute for each individual the predicted probability of being linked. The key identifying assumption is selection on observables: conditional on the included characteristics, linkage is assumed to be as good as random. Under this assumption, inverse-probability weighting recovers the population distribution of the characteristics included in the propensity

model in expectation. Intuitively, individuals who are unlikely to be linked but nevertheless appear in the linked sample receive higher weight, while individuals who are very likely to be linked receive lower weight.

I construct stabilized inverse propensity weights to improve finite-sample performance. Specifically, linked individuals are assigned a weight equal to the ratio of the unconditional linkage rate in the census to their predicted linkage probability. This normalization ensures that the weighted linked sample reproduces the size of the linked population while matching the covariate distribution of the full census. Individuals who are not linked are dropped after the weights are constructed; the weights are only applied to the linked sample in subsequent analysis.

This reweighting step is performed separately for each census year to allow the linkage process and its correlates to vary over time. After reweighting, all family-level and individual-level moments used in estimation and counterfactual analysis are computed using these weights. The weighted linked sample therefore matches the full census more closely along the observed dimensions included in the propensity model.

The main strength of this approach is transparency and flexibility. It directly addresses selection into linkage using a large set of observables that capture demographic composition, geography and economic structure. It also nests simpler reweighting schemes—such as weighting by age or region alone—as special cases. It does not correct for selection on unobserved characteristics, so I interpret the weights as balancing the linked sample on observables rather than fully restoring national representativeness.

A1.2 Internal validation

This appendix assesses the internal reliability of the reconstructed kin network using the census records, the Census Tree crosswalks and the reconstructed lineage files. The exercises validate three distinct stages of the construction. First, I test whether alternative cross-census linking paths identify the same individual. Second, I test whether relationships inferred from the reconstructed network agree with relationships independently observed in a later census. Third, I examine how reliability changes with genealogical distance and the route through which lineage identifiers propagate. The last exercise disciplines the cleaning rules described in Appendix [A1.1](#).

Cross-census identity links. I first test the identity links underlying the network. Whenever both a chained and a direct longer-horizon Census Tree link exist, the two routes should identify the same individual. I compare the chain $1900 \rightarrow 1910 \rightarrow 1920$ with the direct $1900 \rightarrow 1920$ link on a deterministic one-in-sixteen sample of 1900 identifiers. Among individuals for whom both routes are observed, the two links identify the same 1920 individual in 99.67 percent of cases.

Table A1.2.1 shows similarly high agreement across observable characteristics. Differences by sex, race, nativity, urban status and region are small. The identity links therefore exhibit high internal consistency both overall and across the main demographic groups used in the analysis.

Table A1.2.1: Transitive consistency of cross-census identity links

Dimension	Group	Pairs	Agreement
All		1,475,252	99.67%
Sex	Male	783,728	99.50%
	Female	691,524	99.86%
Race	White	1,429,813	99.68%
	Non-white	45,439	99.38%
Residence	Rural	983,545	99.69%
	Urban	491,707	99.63%
Nativity	U.S.-born	1,345,120	99.67%
	Foreign-born	130,132	99.63%
Region	Midwest	620,399	99.70%
	Northeast	386,843	99.65%
	South	388,883	99.65%
	West	79,127	99.65%

Notes: The chained 1900 – 1910 – 1920 link is compared with the direct 1900 – 1920 Census Tree link on a deterministic one-in-sixteen sample of 1900 identifiers. The sample contains individuals for whom both routes exist; the absence of a direct long-horizon link is not treated as a disagreement. Characteristics are measured in 1900.

Relationship holdout. I next test relationships inferred by the network against information from a later census that is not used to form the prediction. For a target census t_2 , I take the reconstructed network as of the preceding wave t_1 and carry it forward using only the $t_1 \rightarrow t_2$ Census Tree link. The household structure observed in t_2 is then used to evaluate the predicted relationship. Where available, I use the census mother, father and spouse pointers; otherwise I use the relationship-pair crosswalk described in Appendix A1.1.

I conduct the exercise for 1900, 1910, 1920 and 1940 using a one-in-twenty systematic sample of households in each target census. The validation unit is a pair of co-residents who already shared a lineage identifier before the target wave.

Because two individuals can share several lineage identifiers, I de-duplicate the validation sample in the same way as the network measures used in the analysis: each unordered pair enters once through the lineage implying the closest genealogical relationship. This yields 3.46 million distinct pairs from 5.66 million individual-by-lineage records.

The relationship prediction is formed before examining the target-wave household and depends on the generation difference and distance between the two reconstructed tree positions. In-law

categories that cannot be distinguished from their blood counterparts in the multigenerational path are combined when scoring agreement. Pairs for which either the census or the reconstructed network does not identify a unique relationship are excluded from the agreement calculation rather than treated as errors.

Agreement across waves. Table A1.2.2 reports the holdout results. Across the four target waves, the reconstructed relationship agrees exactly with the later census for 96.95 percent of verifiable pairs and at the broader close- versus extended-kin level for 98.83 percent. Exact agreement exceeds 96.7 percent in 1910, 1920 and 1940 and is somewhat lower in 1900, when the reconstructed network is based on fewer preceding census waves.

De-duplicating to the closest lineage improves agreement in every target wave. The gain ranges from about 0.3 percentage points in 1910 to 1.3 points in 1900, consistent with redundant, more distant lineage paths being less informative than the closest genealogical connection.

Table A1.2.2: Out-of-sample relationship agreement

Target wave	Pairs	Verifiable	Agreement	Broad class
Closest lineage				
1900 (from 1880)	255,010	239,605	95.50%	98.56%
1910 (from 1900)	958,204	925,524	97.38%	99.00%
1920 (from 1910)	1,033,095	989,467	97.11%	98.97%
1940 (from 1930)	1,210,862	1,159,247	96.76%	98.84%
Pooled	3,457,171	3,313,843	96.95%	98.83%
Individual-by-lineage records				
1900	338,685	318,703	94.25%	98.34%
1910	1,482,188	1,439,117	97.05%	98.92%
1920	1,638,711	1,578,776	96.69%	98.78%
1940	2,195,678	2,119,840	95.93%	98.53%
Pooled	5,655,262	5,456,436	96.35%	98.69%

Notes: Relationships are predicted from the reconstructed network as of the preceding wave and carried forward using the Census Tree link only. No target-wave household information enters the prediction. Verifiable pairs are those for which the target census identifies a relationship and the network assigns a kin category. Broad class groups parent-child, sibling and spouse ties as close kin and grand-parent, aunt/uncle and cousin ties as extended kin. The sample is a one-in-twenty systematic household sample in each target census.

Precision and recall by relationship. Overall agreement can conceal errors in rare relationship categories. Table A1.2.3 therefore reports precision, the fraction of predicted relationships confirmed by the census and recall, the fraction of relationships observed in the census recovered by the network.

Close relationships are measured accurately in both dimensions. Before the final distant-kin cleaning, precision is 98.73 percent for parent–child links, 98.25 percent for siblings and 95.06 percent for spouses, with recall above 94 percent for all three. The cleaning rules leave these relationships essentially unchanged.

Extended relationships are more difficult to validate because few extended kin co-reside and can therefore be observed directly in the census. Grandparents, aunts or uncles and cousins together account for less than one percent of the co-resident pair sample. Their validation moments are consequently based on much smaller and selected samples and errors are more concentrated in these categories because they require longer reconstructed paths. The final cleaning substantially raises precision for grandparents and aunts or uncles, while reducing recall by discarding the least reliable distant paths.

Table A1.2.3: Precision and recall by relationship

Relationship	Observed	Before distant-kin cleaning		After distant-kin cleaning	
		Precision	Recall	Precision	Recall
Parent–child	1,872,634	98.73%	95.53%	98.73%	95.53%
Sibling	905,400	98.25%	95.54%	98.47%	95.25%
Spouse	576,503	95.06%	94.87%	95.06%	94.87%
Grandparent	11,778	39.42%	79.39%	68.49%	74.14%
Aunt/uncle	528,763	30.38%	26.59%	66.31%	15.51%
Cousin	764,401	32.99%	33.67%	54.32%	21.95%

Notes: Closest-lineage pairs pooled across the 1900, 1910, 1920 and 1940 target waves. “Observed” is the number of relationships recorded by the target census and is the denominator of recall. All columns apply closest-lineage de-duplication. The final two columns additionally restrict distant relationships to direct lineage paths and impose the age-plausibility rule described in Appendix A1.1. Ambiguous or unmappable census relationships are excluded from the validation.

Validation using parent pointers. Census parent pointers provide a more direct validation of parent–child and sibling relationships. Among reconstructed parent–child pairs who co-reside, the census pointer confirms the reconstructed father in 97.23 percent of cases and the reconstructed mother in 97.85 percent. Direct contradictions are below one percent for both, while missing pointers are reported separately rather than treated as disagreements.

Sibling relationships show the same pattern. Most reconstructed sibling pairs either share both observed parents or share one parent, as expected for full siblings, half-siblings and households in which only one parent is observed. Direct contradictions are rare. Table A1.2.4 shows similar results across waves; the weaker information in 1900 is driven primarily by missing parent pointers rather than contradictory links.

Table A1.2.4: Validation against census parent pointers

	1900	1910	1920	1940
Parent-child pairs: fathers				
Confirmed	95.57%	97.28%	97.50%	97.31%
Contradicted	0.39%	0.46%	0.64%	0.65%
Pointer missing	4.05%	2.26%	1.85%	2.04%
Parent-child pairs: mothers				
Confirmed	96.93%	97.87%	97.98%	97.93%
Contradicted	0.46%	0.55%	0.62%	0.62%
Pointer missing	2.61%	1.59%	1.40%	1.45%
Sibling pairs				
Both parents shared	56.56%	82.30%	82.05%	82.28%
One parent shared	25.28%	13.34%	12.96%	12.33%
Contradicted	0.12%	0.13%	0.11%	0.14%
No usable pointer	17.93%	4.23%	4.83%	5.20%

Notes: Closest-lineage pairs. For parent-child pairs, the older individual is checked against the younger individual's `momloc` or `poploc` according to the parent's sex. For sibling pairs, the two individuals are compared using both parent pointers. Missing pointers are reported separately and are not treated as contradictions. A small residual category of sibling pairs is identified from the household relationship codes rather than usable parent pointers and is omitted from the displayed rows.

Propagation depth and entry route. The principal weakness of the reconstructed network arises when lineage identifiers are propagated through long genealogical paths. Table A1.2.5 reports relationship agreement by the number of tree steps separating a pair and by the route through which the lineage identifier was inherited.

Direct-blood paths remain highly reliable through two steps and retain 87.90 percent agreement at three steps. Spousal, descendant and collateral routes also perform well at two steps, but their accuracy falls substantially at three steps, to between 48 and 58 percent. At four steps, agreement is below one third for every route and at five or more steps it is below 15 percent. The loss of reliability is therefore concentrated in longer propagation paths rather than in the short connections that constitute most of the reconstructed network.

This pattern motivates restricting distant propagation rather than imposing a general restriction on extended kin. The cleaning rules retain short paths, preserve longer direct-blood connections where they remain informative and remove the distant propagation routes for which the validation indicates substantially greater classification error.

Choice of cleaning rules. Table A1.2.6 compares alternative restrictions for distant relationships. Starting from closest-lineage de-duplication, requiring paths of three or more steps to follow the direct-blood route raises overall agreement from 96.95 to 97.56 percent and extended-

Table A1.2.5: Agreement by propagation depth and entry route

Tree steps	Direct blood	Spouse entry	Descendant	Collateral
1	97.33%	96.62%	87.02%	25.95%
2	97.90%	97.41%	95.85%	95.15%
3	87.90%	58.22%	48.24%	51.23%
4	13.15%	30.85%	31.07%	27.76%
≥ 5	14.42%	8.70%	1.40%	3.30%
Pair records	3,167,672	2,353,333	111,092	23,165

Notes: Individual-by-lineage pair records pooled across the four target waves. The route refers to the way the lineage identifier reached the pair. Direct blood means that neither side inherited the identifier through a spousal, descendant or collateral route. Records at zero tree steps cannot be assigned a relationship and are omitted from the agreement rows; the bottom row reports all records by entry route.

kin precision from 45.75 to 50.16 percent. Adding the age-plausibility restriction raises agreement further to 97.71 percent and extended-kin precision to 60.05 percent, while retaining substantially more extended-kin recall than a blanket two-step depth cap. This combination is used in the baseline construction.

The alternatives clarify the trade-off. Restricting all relationships to two steps produces slightly higher overall agreement, 97.79 percent, but lowers extended-kin recall to 38.29 percent and yields slightly lower extended-kin precision than the adopted rule. A generation-only classifier discards information contained in tree position, while requiring confirmation through two separate lineages removes a large share of otherwise well-measured kin links. The adopted rule therefore targets the part of the network for which the holdout exercise identifies a clear loss of reliability without discarding informative extended relationships more generally.

Genealogical consistency and de-duplication. I conclude with two mechanical checks on the reconstructed network. First, I search for distributions inconsistent with a valid genealogy. Such cases are rare. A small fraction of individual-by-lineage records place two individuals on the same reconstructed tree node, imply a generation difference larger than the number of tree steps, or lack a generation index. Age-based contradictions among predicted parent–child and sibling pairs are similarly uncommon. Records for which the relationship cannot be resolved are not used in relationship-specific measures.

Second, I quantify the importance of de-duplicating overlapping lineages. Multiple lineage membership becomes more common as the reconstructed panel deepens because the same pair can become connected through several ancestors. In the pooled validation sample, 64.2 percent of individual-by-lineage pair records belong to a pair reachable through more than one lineage. Treating these lineage records as distinct kin pairs would therefore mechanically inflate kin counts and increasingly do so in later waves. The baseline measures instead retain each

Table A1.2.6: Alternative cleaning rules for distant kin

Rule	Pairs	All pairs		Extended kin	
		Agreement	Broad	Precision	Recall
Closest lineage only	3,457,171	96.95%	98.83%	45.75%	56.42%
Direct blood for $d \geq 3$	3,420,401	97.56%	99.37%	50.16%	49.18%
Depth cap at $d \leq 2$	3,356,049	97.79%	99.56%	59.30%	38.29%
Generation-only classifier	3,457,171	97.44%	99.41%	59.42%	44.65%
Require two lineages	1,434,824	96.77%	99.32%	51.04%	16.37%
Drop unknown-branch paths	3,361,722	96.95%	98.83%	45.76%	54.83%
Age plausibility for $d \geq 3$	3,457,171	97.14%	99.01%	49.76%	55.73%
Generation-only and direct blood	3,420,401	97.62%	99.48%	60.65%	42.07%
Direct blood and age plausibility	3,420,401	97.71%	99.52%	60.05%	48.61%

Notes: Closest-lineage pairs pooled across the four target waves. Extended kin comprise grandparent, aunt/uncle and cousin relationships. Recall uses the fixed set of extended relationships recorded by the target census in the closest-lineage sample, so dropping pairs cannot mechanically improve recall. The final row combines the two distant-kin restrictions used in the baseline construction. Because the same validation sample is used to compare rules, differences across rows are used for rule selection rather than as an independent out-of-sample test.

unordered pair once through the lineage that implies the closest genealogical connection.

A1.3 External benchmarks for the family network measures

In this section I assess the family network measures using external data against three external sources: the Social Connectedness Index of [Bailey, Cao, Kuchler, Stroebel, and Wong \(2018b\)](#), historical migration networks from [Stuart and Taylor \(2021\)](#) and measures of close-knit communities by [Raz \(2025\)](#). For the first comparison, I construct a comparable Family Connectedness Index (FCI) for two locations i, j :

$$FCI_{i,j} = \frac{Fam_{i,j}}{Fam_i \times Fam_j}$$

where $Fam_{i,j}$ is the total number of family members from region i living in j and Fam_c is the population of commuting zone c . Table [A1.3.1](#) shows a positive correlation between the historical family network and the modern Social Connectedness Index.

For the historical comparison, I use the network connectedness index of [Stuart and Taylor \(2021\)](#), who construct birth-town networks using census and Duke data for 1916–1936. The network index measures the excess correlation in location choices among migrants from the same birth town relative to nearby towns. Formally, it equals the expected increase in flows from j to k when a randomly chosen migrant makes the same move. I investigate their correlation

with a measure of connectedness, controlling for migration distance:

$$\text{Share FM}_{i,j} = \frac{Fam_{i,j}}{Fam_i}$$

Their measure uses origin states and county destinations, so I aggregate county destinations to commuting zones. Table A1.3.1 again shows a positive correlation.³⁵

Table A1.3.1: Correlation of Family Network Measures with Existing Data Sources

VARIABLES	(1) SCI (Bailey et al. 2018)	(2) Network Index (Stuart & Taylor 2021)
FCI (standar.)	0.337*** (0.124)	
Share FM		5.467** (2.269)
Observations	261,002	2,818
R-squared	0.256	0.135

Notes: The unit of observation is a commuting-zone pair. *** denotes significance at the 1 percent level and ** at the 5 percent level. Regressions include log distance between CZs as a control.

I further assess my data with measures of close-knit communities compiled by Raz (2025). He compiles measures constructed from full-count U.S. censuses using first names of individuals. The Local Name Index (LNI), adapted from Fryer and Levitt (2004) is given first name n , locality l , gender g and census year t

$$LNI_{nlg} = 100 \times \frac{\Pr(n \mid l, g, t)}{\Pr(n \mid l, g, t) + \Pr(n \mid -l, g, t)}.$$

Names that are disproportionately common in a locality receive high scores, while names that are common elsewhere receive low scores. The county-level index is the average score of local children.

I construct an analogous statistic using traced family lineages rather than children's first names. For each traced lineage f and commuting zone l , I define

$$FAM_LNI_{fl} = 100 \times \frac{\Pr(f \mid l)}{\Pr(f \mid l) + \Pr(f \mid -l)}.$$

Here $\Pr(f \mid l)$ is lineage f 's population-weighted share of the traced population in commuting zone l and $\Pr(f \mid -l)$ is its share among traced individuals elsewhere in the country. Each indi-

³⁵I will validate these measures more with recent social connectedness variables from Raz (2025).

vidual is assigned the score of their own lineage–commuting-zone cell and the CZ-level measure is the population-weighted average across residents. A high value therefore identifies places whose residents disproportionately belong to lineages that remain geographically concentrated there. Unlike names, lineage membership is not chosen anew each generation, so the measure is constructed using all traced family members rather than only children.

I then compare the family-network LNI the LNI index computed by [Raz \(2025\)](#) as well as other measures of close-knit communities such as intra-community marriage, tightness of family norms, religious homogeneity, kinship propinquity and a summary strength-of-family-ties index. All measures are originally defined at the county level. I aggregate them to commuting zones using population weights the crosswalk files by [Eckert, Gvirtz, Liang, and Peters \(2020\)](#).

Table A1.3.2: Family-network concentration and historical measures of close-knit communities

	Family-network LNI	R^2	Observations
<i>Panel A. Community structure and norms</i>			
Local Name Index	2.375*** (0.364)	0.314	722
Intra-Community Marriage	0.016 (0.013)	0.007	722
Tight Norms Index	0.165*** (0.040)	0.050	722
Religious Homogeneity Index	0.204*** (0.062)	0.056	722
<i>Panel B. Kinship and family ties</i>			
Median Kinship Propinquity	0.038*** (0.008)	0.164	722
Kinship Propinquity Rate	0.016*** (0.003)	0.049	722
Strength of Family Ties Index	0.164*** (0.047)	0.031	722

Notes: The table relates family-network concentration to historical measures of close-knit communities from [Raz \(2025\)](#). Each row reports a separate commuting-zone-level regression. The dependent variable is the measure listed in the first column, aggregated from counties to commuting zones using population weights. The explanatory variable is the family-network Local Name Index (LNI). Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

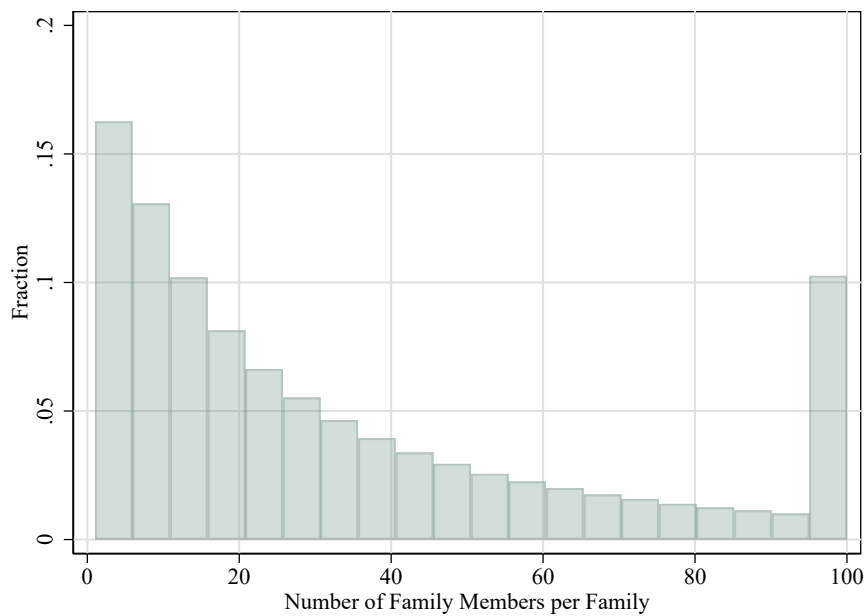
A1.4 Summary statistics of families

Table A1.4.1: Summary statistics of family characteristics in 1940

Characteristic	25th Percentile	Median	75th Percentile
Family Size (No. of Members)	9	22	51
Generational Depth (Avg.)	2.8	3.2	3.5
Number of families per Individ.	1	2	3

Notes: This table reports summary statistics at the family level. Family size refers to the total number of individuals linked to a focal person. Generational depth is the average number of generations represented in a family calculated from the initial household head in 1850.

Figure A1.4.1: Distribution of family members



Notes: This figure shows the distribution of family sizes in 1940 from the final dataset.

Table A1.4.2: Summary statistics: full census vs. linked family-ties samples

	Full census	Ties (unweighted)	Ties (weighted)	Unw. – Census	Wtd. – Census
Observations (Millions)	132	32.2	32.2		
<i>Panel A: Demographics and residence</i>					
Age (years)	31.05	30.92	31.37	-0.13	0.31
Male (%)	50.14	50.89	49.60	0.75	-0.54
White (%)	89.80	99.18	88.42	9.38	-1.38
Married (%)	45.85	48.75	45.97	2.90	0.12
Urban (%)	55.58	42.03	56.86	-13.56	1.27
Homeownership (%)	44.14	50.29	47.41	6.15	3.27
<i>Panel B: Employment and sector</i>					
Agriculture employment (%)	18.39	24.01	16.47	5.62	-1.92
Manufacturing employment (%)	34.11	32.27	33.25	-1.84	-0.86
Self-employed (%)	19.27	25.25	18.94	5.98	-0.33
Farm resident (%)	23.08	31.19	21.97	8.11	-1.11
<i>Panel C: Mobility</i>					
Moved across states (%)	5.40	5.71	5.61	0.30	0.21
Moved within state (%)	53.24	51.16	52.53	-2.08	-0.71
<i>Panel D: Monetary outcomes</i>					
Income (real \$)	44199	42295	45271	-1904	1072
Rent value (real \$)	66.25	68.34	63.26	2.09	-2.99

Notes: Entries are sample means (pooled across the census waves used in the analysis). “Ties” samples consist of individuals for whom extended-family links can be constructed. The weighted ties column applies inverse-propensity weights for linkage so that the linked sample matches the full census on observed characteristics. Percentage variables are reported in percent. Income and rent values are reported in the same real units as used in the main analysis.

Table A1.4.3: Summary statistics by individual characteristics

Variable	Subgroup	FM in Origin	FM not in Origin	Avg. Distance (km)	Avg. Mig. Rate	Direct Mig. Rate	Direct Mig. Share (%)
Race	Non-White	44.44	19.67	543.95	20.52	6.62	32.29
	White	18.77	13.74	542.92	15.58	6.28	40.32
Sex	Female	16.49	10.91	642.76	10.19	4.19	41.09
	Male	19.01	13.87	538.91	15.93	6.41	40.25
Age	> 30	18.56	13.35	549.85	13.65	5.41	39.61
	16–30	20.13	14.67	526.29	23.62	9.89	41.89
Marital Status	Not Married	16.31	11.53	654.16	11.60	4.59	39.57
	Married	19.11	13.91	535.23	15.97	6.44	40.33
College	No	19.80	14.17	519.66	14.60	6.22	42.61
	Yes	11.80	11.01	665.30	23.00	6.68	29.05
Sector	Agriculture	21.88	15.34	423.06	13.05	6.44	49.33
	Manufacturing	20.25	14.68	548.73	16.80	6.82	40.60
	Services	16.21	12.48	571.36	17.35	6.01	34.66
	Not Employed	16.39	12.42	712.12	11.81	4.72	40.01
Born In-State	No	10.92	13.02	679.77	22.85	8.36	36.57
	Yes	22.93	14.35	409.75	11.89	5.22	43.94
Parent	No	15.36	12.61	619.59	16.04	6.24	38.91
	Yes	20.03	14.14	516.46	15.45	6.30	40.76
Homeownership	No	18.35	13.43	533.54	24.45	9.39	38.39
	Yes	19.29	14.42	548.10	9.21	4.15	45.08
Income	Below Median	20.89	14.79	549.45	17.98	7.94	44.18
	Above Median	15.60	11.65	571.88	18.26	5.23	28.65

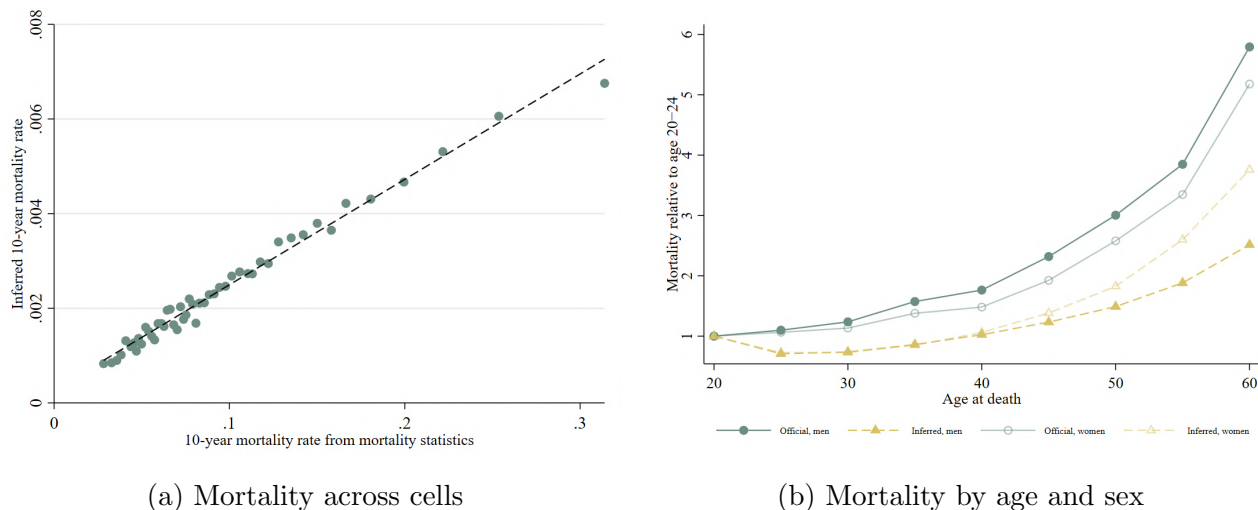
Notes: The table reports subgroup-level averages of family members (FM), migration behavior, and distances. Distances are measured in kilometers. Direct migration share refers to the share of migrants who move directly rather than sequentially. All values are rounded to two decimal places.

A1.5 External benchmark for inferred deaths

Deaths are inferred when an individual who is married in census t reports being widowed in census $t + 10$. I validate this measure against official mortality statistics from the U.S. death-registration area, digitized from the annual *Mortality Statistics* volumes of the U.S. Bureau of the Census by [Miller \(2006\)](#). Because the inferred measure requires marriage, widowhood and successful linkage across censuses, its level need not coincide with population mortality. I therefore compare variation across state, period, sex and age cells and the age profile of mortality.

Figure A1.5.1 shows that the inferred measure closely tracks these independent mortality patterns. Across 2,169 state-by-period-by-sex-by-age cells, the population-weighted correlation between inferred and official mortality is 0.838. Panel (b) shows that the inferred deaths also reproduce the increase in mortality with age for both men and women, although the gradient is flatter at older ages.

Figure A1.5.1: Validation of inferred deaths



Notes: Panel (a) compares inferred ten-year mortality with official mortality across state-by-period-by-sex-by-age cells. Points are weighted by the population eligible to generate an inferred death and the dashed line shows the weighted linear fit. Panel (b) compares age profiles separately by sex, normalizing mortality to one at ages 20–24. Official mortality comes from the historical death-registration statistics. Inferred deaths are identified from transitions from married to widowed status across consecutive censuses.

A1.6 Summary statistics of death variables

Table A1.6.1: Summary Statistics For Death Variables

	Dead individuals (1)	Dead < 40 (2)	Married (census) (3)
Observations	73,330	21,629	49,730,378
<i>Panel A: Demographics</i>			
Age (years)	46.50	32.39	41.07
Male (share)	0.70	0.68	0.50
White (share)	0.89	0.84	0.91
<i>Panel B: Sector and residence</i>			
Agriculture (share)	0.22	0.15	.25
Manufacturing (share)	0.35	0.37	0.34
Services (share)	0.44	0.47	0.4
Farm resident (share)	0.19	0.16	0.24
Homeowner (share)	0.54	0.38	0.48
In labor force (share)	0.70	0.72	0.53
<i>Panel C: Outcomes</i>			
Occupation score	25.36	25.33	24.29
House value	5,488.79	4,399.91	6,366.03
Rent value	40.65	31.78	47.50
Literate (share)	0.97	0.98	0.95

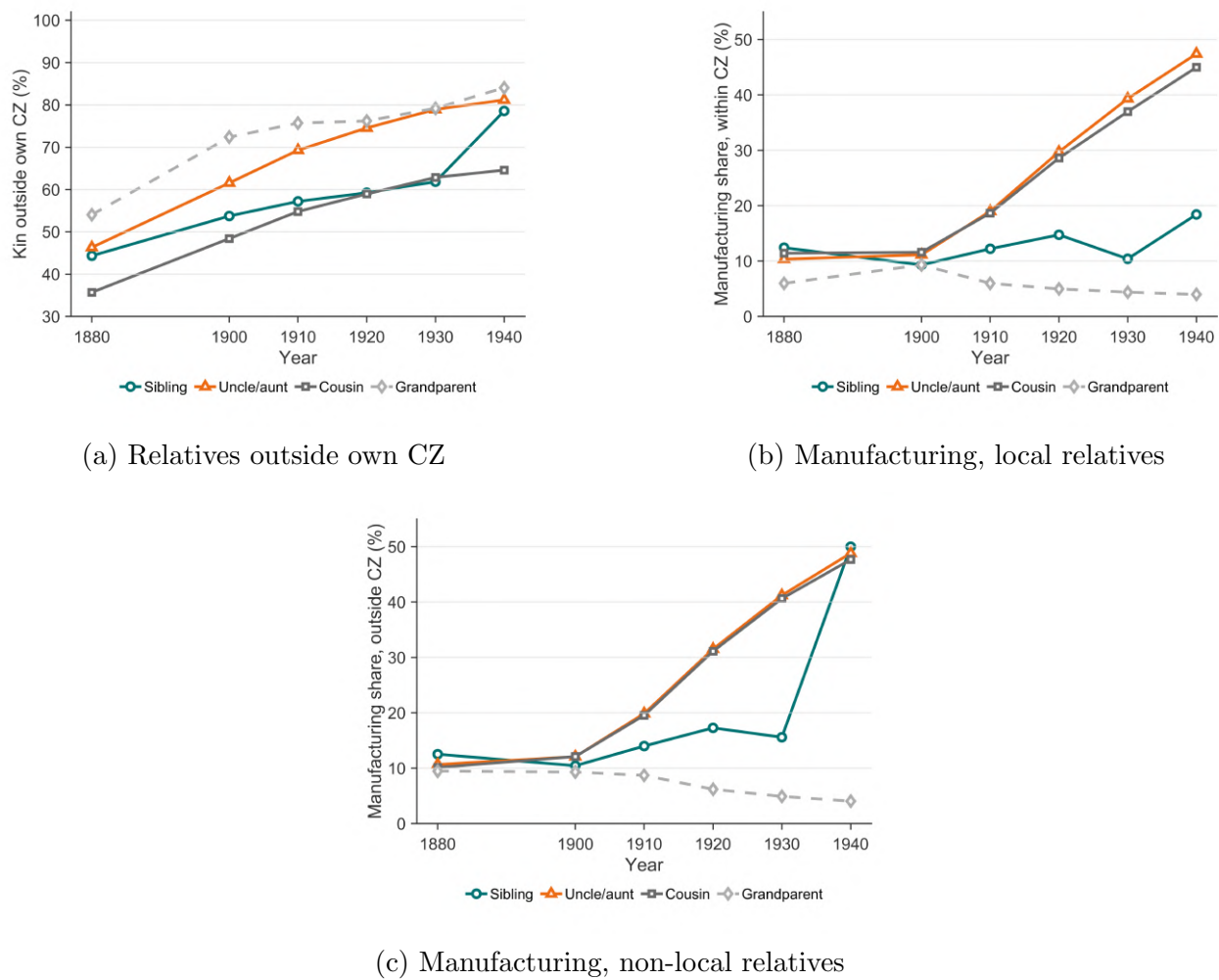
Notes: Entries report sample means. “Dead individuals” are defined by as people whose spouse became widowed between 1930-1940; “Dead < 40” “Dead individuals” that are below 40 years old. Married individuals are defined by having having a spouse present in the household (`marst=1`). Additional sample counts: number of married individuals linkable to the ties data = 10,533,803; number of constructed ties = 31,411,504. Variables measured as shares take values in $[0, 1]$.

A2 Empirics

A2.1 The evolution of family networks across sectors and space

This appendix provides additional descriptive evidence on how family networks changed during structural transformation. The main text shows that cross-CZ family connections became both more geographically dispersed and more oriented toward manufacturing. Here I document that these patterns are visible across relationship types and throughout the distance distribution.

Figure A2.1.1: Family networks across sectors and space

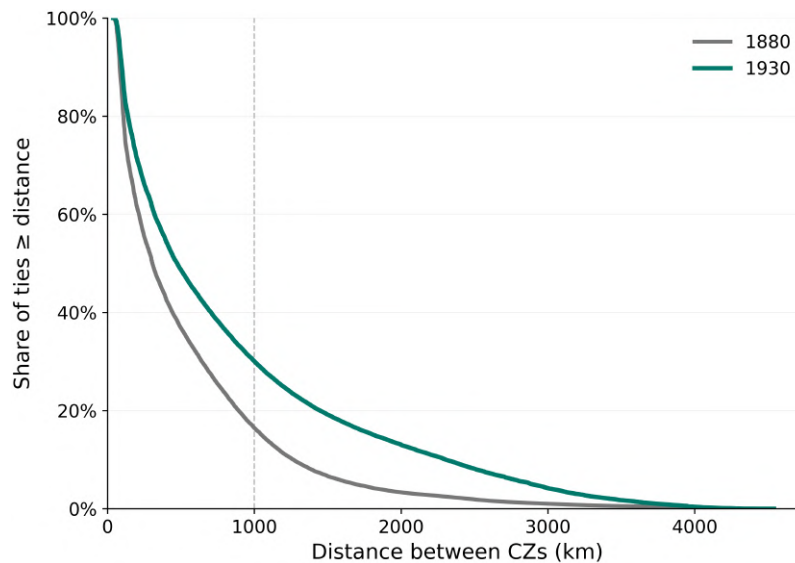


Notes: Panel (a) reports the share of relatives living outside the individual's CZ by relationship and census year. Panels (b) and (c) report the manufacturing share among relatives employed in agriculture or manufacturing, separately for relatives living within and outside the individual's CZ.

Panel (a) shows a broad increase in the geographic dispersion of family networks. By 1940, roughly 80% of aunts and uncles and grandparents and about two-thirds of cousins live outside the individual's CZ. Panels (b) and (c) show a simultaneous shift in sectoral composition. Manufacturing becomes increasingly prevalent among aunts and uncles and cousins both within

and outside the individual's CZ, while the pattern is weaker for siblings and grandparents.

Figure A2.1.2: Spatial dispersion of cross-CZ family ties



Notes: Share of cross-CZ family ties spanning at least a given distance in 1880 and 1930. Distances are measured between CZ centroids.

The increase in geographic dispersion is not driven only by a few very long-distance connections. Among cross-CZ ties, mean distance rises from 542 km in 1880 to 849 km in 1930, the median from 306 to 475 km and the 90th percentile from 1,266 to 2,303 km. The share extending more than 1,000 km increases from 16.5% to 30.1%. Together, these patterns show that structural transformation coincided with a broad reorientation of family networks across both sectors and space.

A2.2 Kinship and place-based social connections

The main analysis focuses on family connections, but kinship is not the only social relationship that can shape economic choices. Workers from the same hometown may exchange information, help one another find jobs or follow similar migration routes even when they are not related. Migrants from the same state or country of birth may likewise concentrate in particular destinations. A natural concern is therefore that the predictive power of kinship simply reflects these broader place-based connections.

I examine this possibility by asking whether kinship predicts sector and migration choices among workers who share the same hometown and birthplace. The exercise is complementary to the quasi-experimental evidence in the main text. It is not intended to identify the causal effect of kinship relative to other forms of social connection. Instead, it asks whether family ties contain information about workers' choices beyond the broader origin-based connections that they share with other people from the same place.

Migration. I use linked individuals in the full-count censuses from 1900 to 1940. For each census year $t \in \{1910, 1920, 1930\}$, I select linked individuals aged 18–45 and observe their location ten years later. For computational tractability, I take a 2% random sample of individuals and pair each person with each of the 40 largest commuting zones (CZs) other than the person’s current CZ. A person’s hometown is the county in which they live at t . Birthplace is the state of birth for the U.S.-born and country of birth for the foreign-born, since the census does not report town of birth.

For person i , origin county $c(i)$, birthplace $b(i)$ and potential destination d , I estimate

$$\text{Move}_{idt} = \beta \text{Kin}_{idt} + \alpha_{c(i)dt} + \gamma_{b(i)dt} + \varepsilon_{idt}, \quad (\text{A2.2.1})$$

where $\text{Move}_{idt} = 100$ if person i lives in destination d at $t + 10$. The variable Kin_{idt} equals one if a sibling, grandparent, aunt or uncle, or cousin lives in d at time t .

The fixed effects absorb two broad sources of place-based social connection. Origin-county $\times$ destination $\times$ year fixed effects absorb anything shared by people currently living in the same county that makes a particular destination attractive. This includes established migration routes from the hometown: for example, the share of residents of county c at $t - 10$ who live in d at t . Birthplace $\times$ destination $\times$ year fixed effects analogously absorb destination-specific connections shared by people born in the same state or foreign country. The coefficient β is therefore identified from differences in whether individuals have relatives at a destination after absorbing the migration patterns associated with both their current hometown and their birthplace.

Family connections remain strongly related to destination choice. With only origin-county and destination fixed effects, having kin in a destination raises the probability of moving there by 1.74 percentage points. Absorbing the full set of hometown-to-destination and birthplace-to-destination fixed effects reduces the coefficient to 0.95 percentage points, compared with a migration rate of 0.11% for person–destination pairs without kin. Thus, more than half of the original association remains and workers are about ten times as likely to move to a destination where relatives already live.

The result is not driven by a small set of established hometown migration routes. Among destinations to which nobody from the individual’s hometown had moved during the previous decade, kin at the destination still raise migration by 0.52 percentage points, relative to a baseline migration probability of 0.014%. The association is also present for destinations more than 500 kilometers from the origin and among foreign-born workers, for whom the birthplace fixed effects absorb country-of-origin-specific destination patterns. Estimating the effects of hometown and birthplace migration shares directly leads to the same conclusion: both predict destination choice, but adding family ties leaves their coefficients largely unchanged. Kinship therefore captures variation in destination access that is distinct from these broader origin-based

migration patterns.

Sector choice. I use a similar exercise to distinguish sector-specific family connections from broader characteristics shared by workers from the same places. The sample contains a 30% draw of agricultural workers aged 18–45 who are linked to the following census. I estimate

$$\text{Manuf}_{i,t+10} = \theta \text{KinM}_{it} + \lambda \text{KinE}_{it} + \kappa n_{it} + \delta_{c(i)b(i)t} + u_{it}, \quad (\text{A2.2.2})$$

where KinM_{it} indicates that a sibling, grandparent, aunt or uncle works in manufacturing, KinE_{it} indicates that at least one such relative is employed and n_{it} is the number of observed relatives. The specification includes origin-county $\times$ birthplace $\times$ year fixed effects. It therefore compares workers from the same hometown and birthplace while conditioning on whether they have employed relatives and on the number of relatives observed. Standard errors are clustered by origin county.

Having a relative in manufacturing raises the probability that an agricultural worker enters manufacturing over the next decade by 0.48 percentage points, relative to a baseline manufacturing-entry rate of 6.75%. The estimate remains 0.43 percentage points after adding age $\times$ year fixed effects. What matters is not simply having relatives who are economically active. Conditional on manufacturing kin, the indicator for having any employed relative enters negatively. Manufacturing relatives also reduce the probability of remaining in agriculture by 1.10 percentage points, while the corresponding increase in services is smaller and imprecisely estimated. The association is therefore specific to the sector in which relatives work rather than to employment among relatives more generally.

Interpretation. These exercises show that the role of kinship in the main analysis is not simply a proxy for broad migration patterns shared by people from the same hometown or birthplace. Family connections predict both where workers move and which sector they enter after flexibly absorbing these place-based similarities. The distinction is particularly clear for migration: relatives predict moves even toward destinations with no recent migration from the worker’s hometown.

The exercise does not imply that family ties are the only relevant social connections. The historical censuses do not observe friendships, neighbors or other non-kin relationships directly, and unobserved preferences may also be correlated within families. The results instead establish a narrower point that is important for the quantitative analysis: kinship contains substantial destination- and sector-specific information that is not captured by broad origin-based social connections.

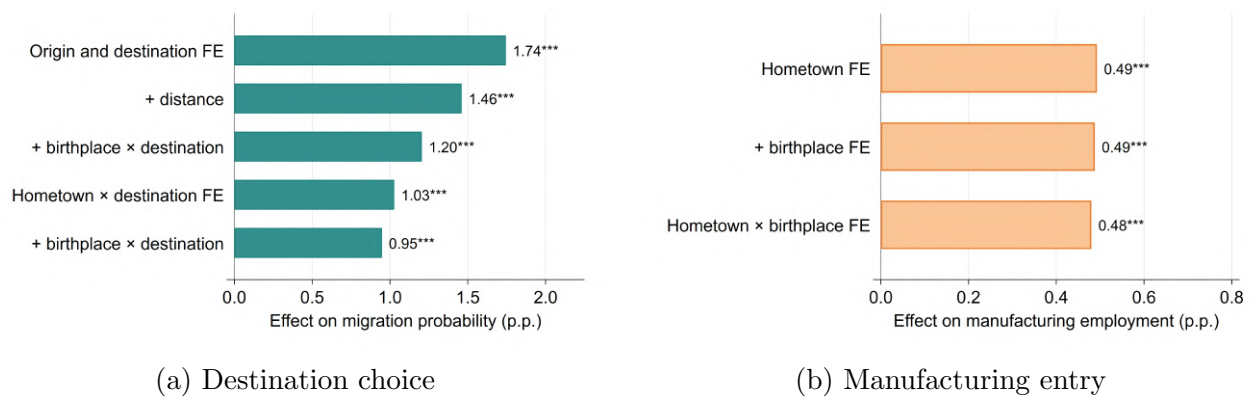


Figure A2.2.1: Kinship beyond hometown and birthplace connections
Notes: Panel (a) reports estimates of β from equation (A2.2.1). Panel (b) reports estimates of θ from equation (A2.2.2). Outcomes are measured in percentage points. Each specification successively adds the fixed effects or controls indicated in the figure. In panel (a), hometown denotes the county of residence at time t and birthplace denotes the state of birth for the U.S.-born and country of birth for the foreign-born. All fixed effects are interacted with census year. Standard errors are clustered by origin county. Stars denote significance at the 10%, 5% and 1% levels.

Table A2.2.1: Kinship beyond place-based connections

	Coefficient	s.e.	Mean, no kin	Observations
Panel A. Destination choice: kin at destination				
All	0.948	(0.024)	0.105	19,231,125
<i>By census year</i>				
1910	1.239	(0.047)	0.107	5,469,044
1920	1.071	(0.041)	0.127	6,370,686
1930	0.706	(0.028)	0.085	7,391,395
<i>By birthplace</i>				
Born in the United States	0.948	(0.024)	0.105	19,004,821
Foreign-born	0.803	(0.174)	0.133	180,489
<i>By destination</i>				
No recent hometown migration	0.519	(0.037)	0.014	7,887,628
Positive recent hometown migration	1.003	(0.027)	0.170	11,342,216
At least 500 km from origin	0.556	(0.023)	0.051	14,846,361
Panel B. Sector choice: kin in manufacturing				
Manufacturing	0.480	(0.106)	6.747	1,325,569
Manufacturing, age × year FE	0.425	(0.106)	6.747	1,325,569
Services	0.235	(0.135)	12.471	1,325,569
Agriculture	−1.104	(0.193)	66.536	1,325,569

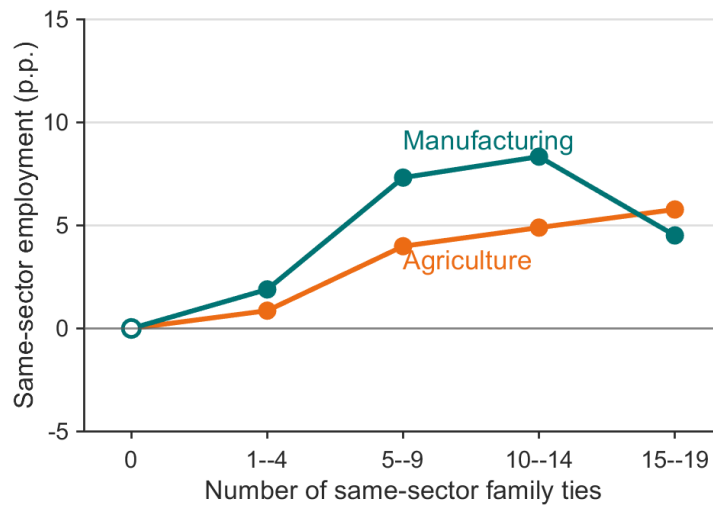
Notes: Outcomes are measured in percentage points. Panel A estimates equation (A2.2.1) with origin county × destination × year and birthplace × destination × year fixed effects on the indicated subsample. “No recent hometown migration” denotes destinations at which no resident of the individual’s origin county at $t - 10$ lives at t . Panel B estimates equation (A2.2.2) with origin county × birthplace × year fixed effects for the indicated sector at $t + 10$. Standard errors are clustered by origin county.

A2.3 Family connections and sectoral choice

A2.3.1 Sectoral concentration in family networks

Figure A2.3.1 reports the conditional relationship between sector employment and the number of same-sector relatives residing in the individual's CZ. The regressions absorb CZ fixed effects and control for individual characteristics and total family size, with individuals who have no same-sector relatives as the omitted category. Sector employment rises strongly with same-sector family presence in both manufacturing and agriculture.

Figure A2.3.1: Sectoral ties and sector choice



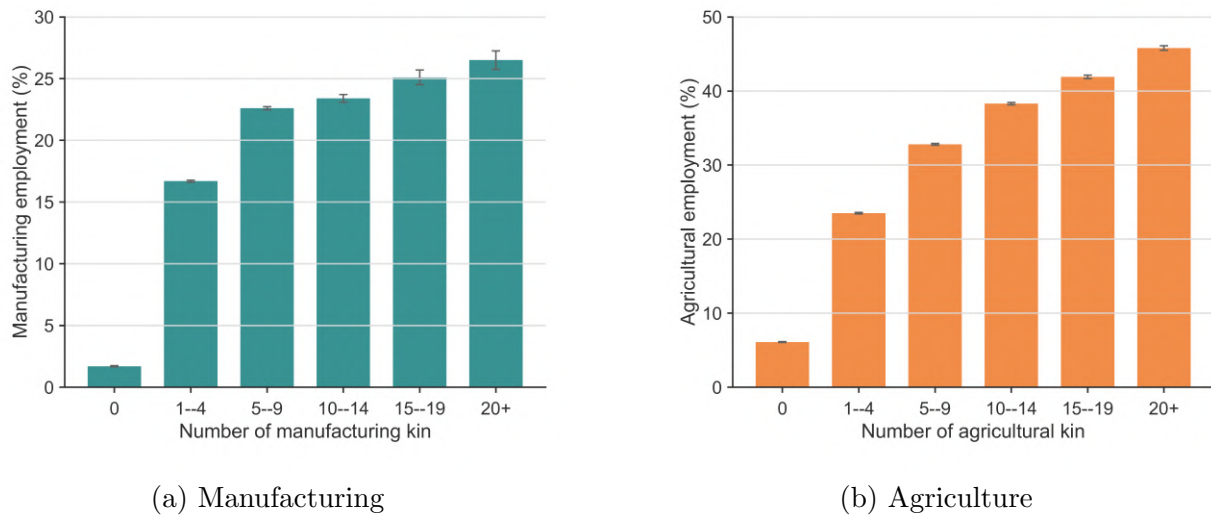
Notes: Coefficients from regressions of sector employment on bins of same-sector relatives residing in the individual's CZ; zero same-sector relatives is the omitted category. Regressions include CZ fixed effects, individual controls and total family size. Vertical bars denote 95% confidence intervals. $N = 9,485,355$.

Figure A2.3.2 reports the corresponding unconditional sectoral employment rates in 1900. The raw gradients are large in both sectors. Manufacturing employment is below 2% among individuals with no local manufacturing relatives and exceeds 20% among individuals with at least five. Agricultural employment rises from roughly 6% among individuals with no local agricultural relatives to more than 30% among those with at least five and exceeds 40% at higher levels of agricultural family presence.

A2.3.2 Deaths of sector-employed relatives: robustness and falsification

The main text focuses on the loss of manufacturing-employed relatives because manufacturing is the central destination sector in the structural-transformation analysis. This section reports

Figure A2.3.2: Sectoral employment and same-sector family ties



Notes: Mean sectoral employment by the number of same-sector relatives residing in the individual's CZ in 1900. Pairwise family relationships are de-duplicated as described in Section 2. Vertical bars denote 95% confidence intervals.

complementary evidence across sectors and examines whether the manufacturing response can be explained by more general consequences of a family death.

Summary statistics This section documents the empirical sample and its characteristics relative to the census.

Alternative sectors. Figure A2.3.3 reports analogous event studies around the deaths of relatives employed in agriculture and services. In each panel, the outcome is employment in the deceased relative's sector. The direction of the response is similar to the manufacturing result: following the loss of a sector-employed relative, subsequent employment in that sector declines. The relationship between sector-specific family connections and sectoral attachment is therefore not unique to manufacturing.

Placebo deaths and specification robustness. A concern with the death design is that a family death may affect subsequent outcomes even when the deceased relative does not provide a sector-specific labor-market connection. I therefore repeat the analysis using deaths of non-employed relatives. Such deaths may remove emotional, financial or household support, but they do not remove a contemporaneous employment connection to manufacturing. The corresponding estimate is close to zero.

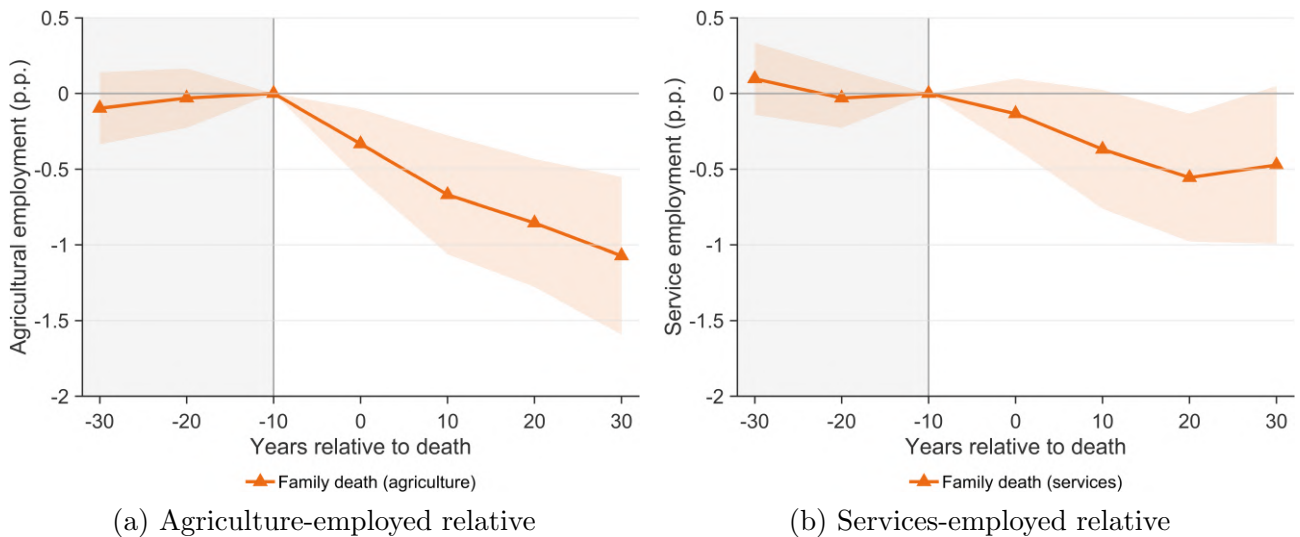
Table A2.3.2 reports this placebo together with alternative specifications of the manufacturing-death design. The estimates are similar when individual and family controls are omitted and under alternative functional forms. The stability of the effect across specifications, together

Table A2.3.1: Empirical sample summary statistics

	Mean		Normalized difference
	Weighted sample	Full-count census	
Age	32.198	33.281	-0.084
Female	0.487	0.490	-0.005
White	0.912	0.894	0.061
Married	0.573	0.562	0.023
Number of children	1.171	1.147	0.015
Owens home	0.547	0.475	0.145
Urban	0.511	0.524	-0.027
Farm residence	0.282	0.267	0.034
In school	0.134	0.113	0.072
In manufacturing	0.102	0.105	-0.009
In agriculture	0.141	0.130	0.034

Notes: The table compares characteristics of the linked panel after post-stratification weighting with the corresponding means in the full-count census. Statistics are measured at the person-year level and pool census waves from 1900 to 1940. Post-stratification weights are the ratio of full-count census to linked-sample observations within cells defined by age band, sex, race, state, urban status and farm residence. The final column reports the normalized difference between the weighted linked sample and the full-count census.

Figure A2.3.3: Deaths of sector-employed relatives in agriculture and services



Notes: The figure reports estimates from the stacked event-study specification in equation (2). Panel (a) reports event-time coefficients around the first eligible death of an agriculture-employed relative and Panel (b) around the first eligible death of a services-employed relative. In each panel, the outcome is employment in the deceased relative's sector. Event time zero is the first census following the death and the omitted period is the preceding census. The specification includes individual-by-stack fixed effects and pre-event-CZ-by-year-by-stack fixed effects, together with age controls and predetermined family and household characteristics. Observations are censored from the second eligible death onward. Regressions use the corrective stack weights together with the population reweighting weights. Standard errors are clustered at the de-duplicated family-identifier level. Vertical bars and shaded regions denote 95% confidence intervals.

with the absence of a comparable response following deaths of non-employed relatives, supports the interpretation that the manufacturing result is linked to the loss of a sector-specific family connection rather than to family bereavement alone.

Table A2.3.2: Sectoral choice and family deaths: robustness

	Dependent variable: manufacturing employment			
	Baseline (1)	No controls (2)	PPMLE (3)	Placebo death (4)
Death (manuf.) –30 years	0.0478 (0.123)	0.0470 (0.123)	-0.0118 (0.0901)	
Death (manuf.) –20 years	-0.0554 (0.132)	-0.0617 (0.133)	-0.0486 (0.0546)	
Death (manuf.)	-0.342** (0.137)	-0.340** (0.139)	0.0558 (0.0352)	
Death (manuf.) +10 years	-0.427** (0.182)	-0.428** (0.183)	-0.1807*** (0.0541)	
Death (manuf.) +20 years	-0.486** (0.233)	-0.404* (0.225)	-0.258*** (0.0670)	
Death (manuf.) +30 years	-1.002*** (0.356)	-0.973*** (0.356)	-0.451*** (0.0940)	
Death (nonemp.) –30 years				0.0811 (0.259)
Death (nonemp.) –20 years				0.168 (0.251)
Death (nonemp.) –10 years				0.0798 (0.241)
Death (nonemp.)				-0.0129 (0.223)
Death (nonemp.) +10 years				0.160 (0.259)
Death (nonemp.) +20 years				0.0924 (0.212)
Observations	22,115,970	22,115,970	3,727,444	22,115,970
R-squared	0.358	0.360		0.358
Mean outcome	36.85	36.85	36.85	36.85
Mean manuf./agric. death	.02	.02	.02	0
Individual FE	✓	✓	✓	✓
CZ-year FE	✓	✓	✓	✓
Individual controls	✓	×	✓	✓
Population weights	✓	✓	✓	✓
Estimation	OLS	OLS	PPMLE	OLS

Notes: The table reports robustness exercises for the effect of losing a manufacturing-employed relative on subsequent manufacturing employment. The baseline stacked specification includes individual-by-stack fixed effects and pre-event-CZ-by-year-by-stack fixed effects, together with age controls and predetermined family and household characteristics. Observations are censored from the second eligible death onward. Regressions use the corrective stack weights together with the population reweighting weights. Standard errors, reported in parentheses, are clustered at the de-duplicated family-identifier level. Column (2) omits the individual and family controls. Column (3) estimates the corresponding alternative functional-form specification. Column (4) replaces deaths of manufacturing-employed relatives with deaths of non-employed relatives as a placebo. The sample covers census observations from 1900 to 1940. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Other outcomes around the death. I next examine whether the manufacturing response coincides with changes in other household outcomes that could indicate a broader disruption following the death. Table A2.3.3 reports responses in employment, household-head status,

number of children, marital status and farm residence. For manufacturing deaths, there is little evidence of corresponding discontinuities in these outcomes. This makes a broad change in household composition or labor-force attachment less consistent with the manufacturing result. The agricultural case warrants somewhat more caution. Deaths of agriculture-employed relatives are accompanied by a decline in farm ownership or farm residence in some specifications. Because the farm is both a productive asset and a place of employment, agricultural deaths may affect sectoral outcomes through succession, inheritance or asset reallocation in addition to the loss of a sector-specific labor-market connection.

Additional heterogeneity. Table A2.3.4 reports the manufacturing response separately by gender and race. The estimates are broadly similar across groups.

Robustness of the stacked event-study design This appendix examines the sensitivity of the manufacturing death event study to treatment-cohort composition, event-time support, the choice of comparison group and population weighting. I estimate the event study around the first eligible death of a manufacturing-employed relative. Because deaths are observed only between censuses, a relative observed alive in census year y and inferred to have died before census $y + 10$ is treated as dying during the interval $(y, y + 10]$. Event time 0 is the first census after the death and the final census before the death, $t = -10$, is the omitted period. The estimation panel is decadal and spans 1900–1940, with individuals observed in all five census waves. The resulting treatment cohorts are 1910, 1920, 1930 and 1940. Within each stack, treated individuals are compared with never-treated individuals and with members of later cohorts while they remain untreated; observations after an individual’s own treatment are not used as controls. Individuals remain in the sample after a first eligible death until a second eligible death occurs, at which point subsequent observations are censored. The baseline specification includes individual-by-stack fixed effects and pre-event-CZ-by-year-by-stack fixed effects. The latter hold location fixed at its pre-event value, thereby absorbing changes in the initial local labor market without conditioning on subsequent migration. Controls include age and predetermined family and household characteristics interacted with census-year effects. Corrective stacked-event-study weights account for the repeated use of controls across stacks and are combined with the population weights used to make the linked sample representative of the full census.

Table A2.3.5 varies features of this design one at a time. The *unweighted* specification removes the population weights while retaining the corrective stacked weights. The *balanced post window* keeps only cohorts observed through +30; because the panel ends in 1940, this leaves the 1910 treatment cohort alone. The *balanced pre-trend window* fixes cohort composition over the pre-treatment window through $t = -20$. The *never-treated controls* specification removes later-treated individuals from the comparison group altogether. Finally, the *fully balanced individuals*

Table A2.3.3: Demographic outcomes and family deaths

	Dependent variable				
	Employed (1)	House (2)	Number of children (3)	Married (4)	Farm (5)
Death (agric.) –30 years					-0.001 (0.036)
Death (agric.) –20 years					0.003 (0.026)
Death (agric.) –10 years					0.001 (0.017)
Death (agric.)					0.008 (0.011)
Death (agric.) +10 years					-0.097*** (0.014)
Death (agric.) +20 years					-0.125*** (0.014)
Death (agric.) +30 years					-0.142*** (0.016)
Death (manuf.) –30 years	-0.002 (0.036)	-0.010 (0.036)	-0.184*** (0.059)	-0.041* (0.024)	
Death (manuf.) –20 years	0.013 (0.026)	-0.020 (0.032)	-0.053 (0.034)	-0.018 (0.017)	
Death (manuf.)	-0.009 (0.024)	-0.005 (0.025)	-0.040 (0.040)	-0.032** (0.013)	
Death (manuf.) +10 years	0.003 (0.029)	0.012 (0.022)	-0.161** (0.060)	-0.033 (0.021)	
Death (manuf.) +20 years	-0.023 (0.033)	0.041 (0.033)	-0.090 (0.093)	-0.001 (0.020)	
Death (manuf.) +30 years	-0.033 (0.043)	0.046 (0.050)	-0.186 (0.148)	-0.024 (0.037)	
Observations	22,115,970	21,789,776	21,845,180	21,845,180	21,845,180
R-squared	0.560	0.457	0.638	0.694	0.580
Mean outcome	0.63	0.52	0.89	0.45	0.36
Mean manuf./agric. death	0.02	0.02	0.02	0.02	0.02
Individual FE	✓	✓	✓	✓	✓
CZ-year FE	✓	✓	✓	✓	✓
Individual controls	✓	✓	✓	✓	✓
Population weights	✓	✓	✓	✓	✓
Estimation	OLS	OLS	OLS	OLS	OLS

Notes: The table reports responses of demographic and household outcomes to the first eligible death of a manufacturing-employed relative using the stacked event-study specification in equation (2). Columns consider employment, household-head status, number of children, marital status and farm residence. All specifications include individual-by-stack fixed effects and pre-event-CZ-by-year-by-stack fixed effects, together with age controls and predetermined family and household characteristics. Observations are censored from the second eligible death onward. Regressions use the corrective stack weights together with the population reweighting weights. Standard errors, reported in parentheses, are clustered at the duplicated family-identifier level. The sample covers census observations from 1900 to 1940. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

specification requires complete observation over every census needed for the corresponding stack and event window. All other features of the baseline specification are held fixed within each exercise.

Event-time support is necessarily asymmetric in a finite panel. Observations close to treatment draw on all four cohorts, whereas longer leads and lags are identified by progressively fewer cohorts. In particular, +10 is supported by the 1910–1930 cohorts, +20 by 1910 and 1920

Table A2.3.4: Sectoral choice and family deaths: heterogeneity

	Dependent variable: manufacturing employment				
	All (1)	Men (2)	Women (3)	White (4)	Non-white (5)
Death (manuf.) –30 years	0.0478 (0.123)	0.0612 (0.127)	0.388 (0.370)	0.0809 (0.119)	-1.610 (3.247)
Death (manuf.) –20 years	-0.0554 (0.132)	-0.0284 (0.147)	0.0222 (0.296)	-0.0181 (0.127)	-5.131 (3.793)
Death (manuf.)	-0.342** (0.137)	-0.343** (0.150)	-0.140 (0.313)	-0.341*** (0.124)	2.482 (3.036)
Death (manuf.) +10 years	-0.427** (0.182)	-0.413** (0.205)	-0.670* (0.377)	-0.423** (0.165)	-4.159 (3.680)
Death (manuf.) +20 years	-0.486** (0.233)	-0.446* (0.250)	-0.533 (0.507)	-0.415* (0.225)	1.011 (3.624)
Death (manuf.) +30 years	-1.002*** (0.356)	-0.945** (0.413)	-0.692 (0.612)	-0.933*** (0.353)	
Observations	22,115,970	18,922,530	3,193,440	22,078,300	37,670
R-squared	0.358	0.394	0.215	0.359	0.578
Mean outcome	36.85	36.85	36.85	36.85	36.85
Mean manuf./agric. death	.02	.02	.02	.02	.02
Individual FE	✓	✓	✓	✓	✓
CZ-year FE	✓	✓	✓	✓	✓
Individual controls	✓	✓	✓	✓	✓
Population weights	✓	✓	✓	✓	✓
Estimation	OLS	OLS	OLS	OLS	OLS

Notes: The table reports heterogeneity in the effect of losing a manufacturing-employed relative on subsequent manufacturing employment using the stacked event-study specification in equation (2). Columns split the sample by gender and race. All specifications include individual-by-stack fixed effects and pre-event-CZ-by-year-by-stack fixed effects, together with age controls and predetermined family and household characteristics. Observations are censored from the second eligible death onward. Regressions use the corrective stack weights together with the population reweighting weights. Standard errors, reported in parentheses, are clustered at the de-duplicated family-identifier level. The sample covers census observations from 1900 to 1940. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

and +30 by the 1910 cohort alone; Table A2.3.6 reports the corresponding sample support. The balanced-window exercises therefore trade comparability of cohort composition against precision and, at the longest horizons, identify effects from substantially narrower treatment populations.

The estimates are stable across the main alternatives. Removing population weights produces almost identical coefficients throughout the event window. Restricting the comparison group to never-treated individuals also leaves the estimates through +20 essentially unchanged, indicating that the baseline results are not driven by the use of not-yet-treated observations as controls. Estimates from balanced windows are less precise because they rely on fewer treatment cohorts, but continue to show a decline in manufacturing employment following the death. In particular, the balanced post specification, which is identified entirely from the 1910 cohort, yields negative estimates at every available post-treatment horizon.

The exercises also clarify how the longest-horizon estimate should be interpreted. The baseline +30 coefficient is identified only from the 1910 treatment cohort and is therefore not directly comparable to shorter-horizon coefficients that average across several cohorts. Its larger mag-

Table A2.3.5: Stacked event study: robustness to cohort composition, controls and panel support

	-30	-20	0	+10	+20	+30
Baseline	0.048 (0.123)	-0.055 (0.132)	-0.342** (0.137)	-0.427** (0.182)	-0.486** (0.233)	-1.002*** (0.356)
Unweighted	0.047 (0.123)	-0.055 (0.134)	-0.327** (0.135)	-0.426** (0.183)	-0.484* (0.284)	-0.982*** (0.374)
Balanced post window	—	—	-0.830*** (0.298)	-0.636 (0.404)	-0.705** (0.331)	-0.721** (0.317)
Balanced pre-trend window	—	0.189 (0.202)	-0.252 (0.199)	-0.316 (0.217)	-0.424* (0.253)	—
Never-treated controls only	0.047 (0.123)	-0.055 (0.134)	-0.327** (0.135)	-0.426** (0.183)	-0.483* (0.284)	-0.538* (0.320)
Fully balanced individuals	-0.015 (0.458)	0.043 (0.327)	-0.497* (0.301)	-0.259 (0.441)	-0.242 (0.311)	-0.361 (0.311)

Notes: The dependent variable is manufacturing employment multiplied by ten; $t = -10$ is omitted. The baseline uses corrective stacked-event-study and population weights. The unweighted row drops the population weights; the remaining rows modify cohort, event-window or comparison-group support as described in the text. Dashes indicate event times outside the corresponding estimation window. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

nitude may reflect either the evolution of the response with time since treatment or differences specific to the earliest cohort. The balanced-post estimate confirms a persistent negative response within that cohort, but is smaller than the baseline +30 estimate. I therefore place greater weight on the common negative post-treatment pattern than on comparisons of coefficient magnitudes across horizons with changing cohort support.

A2.3.3 Occupational entry routes and family persistence

The death design establishes that losing a sector-employed relative lowers subsequent attachment to that sector. I next ask whether the strength of family persistence varies with the availability of alternative routes into an occupation. This provides complementary evidence on the interpretation of sector-specific family ties as a labor-market input.

Classification of occupational entry routes. I classify occupations using the 1949 *Occupational Outlook Handbook* (U.S. Bureau of Labor Statistics, 1949). The Handbook contains separate reports for 209 occupations outside agriculture, including explicit discussions of training and entry for 181 occupations. I classify an occupation as having a *workplace route* when the Handbook identifies apprenticeship, helper or trainee positions, on-the-job training, learning

Table A2.3.6: Treated and control support by event time

Event time	Cohorts included	No. cohorts	Treated individuals	Control ind.
−30	1930, 1940	2	13,668	2,324,726
−20	1920, 1930, 1940	3	16,914	3,492,041
−10	1910, 1920, 1930, 1940	4	18,807	4,660,978
0	1910, 1920, 1930, 1940	4	18,807	4,660,978
+10	1910, 1920, 1930	3	11,203	3,492,041
+20	1910, 1920	2	5,056	2,324,726
+30	1910	1	1,853	1,160,481
Total	—	—	86,308	22,115,970

Notes: The table reports event-time support in the stacked event-study sample. Cohorts are defined by treatment year. The omitted event period is −10.

from experienced workers, promotion from related employment or prior occupational experience as a route into the occupation. I classify an occupation as having a *school route* when the Handbook identifies vocational, trade or technical school, classroom instruction or correspondence instruction as a route of preparation.

I map the Handbook reports to detailed 1950 occupation codes. Because several Handbook reports can map into a single census occupation, I assign a route only when the underlying reports agree; ambiguous codes are excluded from the route-specific coefficients. This procedure yields 33 workplace-only occupations and 10 occupations with a school route. Airplane mechanics remain in the occupational choice set but load on neither route coefficient, producing 44 occupation clusters in the estimation.

Empirical specification. The estimation sample uses occupational outcomes from 1920, 1930 and 1940, with 1900 and 1910 supplying predetermined occupational histories. I retain family-by-generation-by-year cells with at least one predetermined occupational tie somewhere in the choice set, since cells without such a tie contain no within-cell variation in the family-experience measure.

For family f , generation g , year t and occupation o , I estimate

$$E[n_{fgto} \mid \cdot] = \exp\{\beta_W D_{fgto} \mathbf{1}\{\text{workplace only}\}_o + \beta_S D_{fgto} \mathbf{1}\{\text{school route}\}_o + \alpha_{fgt} + \delta_{ot}\}, \quad (\text{A2.3.1})$$

where n_{fgto} is the number of employed family members in occupation o and D_{fgto} indicates that a relative outside the current family-generation-year cell held occupation o in an earlier census. Family-by-generation-by-year fixed effects α_{fgt} absorb the family's overall occupational allocation, while occupation-by-year fixed effects δ_{ot} absorb occupation size and aggregate changes in occupational demand. Identification therefore comes from whether predetermined family ex-

Table A2.3.7: Detailed occupations by route of entry

<i>Panel A. Workplace route only (33 occupations)</i>			
500	Bakers	565	Paperhangers
501	Blacksmiths	575	Pressmen and plate printers
502	Bookbinders	582	Shoemakers and repairers
503	Boilermakers	585	Structural metal workers
505	Cabinetmakers	592	Tool makers and die makers
513	Cranemen, derrickmen, hoistmen	593	Upholsterers
515	Electricians	594	Craftsmen, n.e.c.
520	Electrotypers and stereotypers	621	Attendants, auto service
522	Excavating and road machinery op.	622	Blasters and powdermen
524	Forgemen and hammermen	624	Brakemen, railroad
530	Glaziers	635	Filers, grinders, polishers
533	Inspectors, n.e.c.	641	Furnacemen and smeltermen
535	Job setters, metal	642	Heaters, metal
541	Locomotive engineers	643	Laundry and dry cleaning op.
542	Locomotive firemen	670	Painters, except construction
560	Millwrights	681	Switchmen, railroad
563	Opticians and lens grinders		
<i>Panel B. School route also available (10 occupations)</i>			
504	Brickmasons and stonemasons	573	Plasterers
510	Carpenters	574	Plumbers and pipe fitters
525	Furriers	591	Tinsmiths and sheet metal workers
550	Automobile mechanics	644	Meat cutters
553	Railroad and car shop mechanics		
564	Painters, construction		

Notes: Detailed OCC1950 occupations for which the underlying Handbook reports agree on the route of entry. Codes with conflicting classifications are excluded. Airplane mechanics (545) remain in the occupational choice set but are assigned to neither route.

perience disproportionately predicts entry into particular occupations within the same family-generation-year choice set.

The final estimation sample contains approximately 16.3 million occupation-choice observations from 225,942 reconstructed family identifiers. Because the route classification varies at the occupation level, standard errors are two-way clustered by family identifier and detailed occupation.

Results. Figure A2.3.4 shows a pronounced difference across entry routes. The estimated family-persistence coefficient is $\hat{\beta}_W = 1.051$ for workplace-only occupations and $\hat{\beta}_S = 0.548$ for occupations in which a school route is also available. The difference, $\hat{\beta}_W - \hat{\beta}_S = 0.503$, is statistically significant ($p = 0.005$). In proportional terms, predetermined family experience is therefore substantially more predictive of occupational entry when preparation occurs primarily

Table A2.3.8: Examples of occupational entry routes in the 1949 Handbook

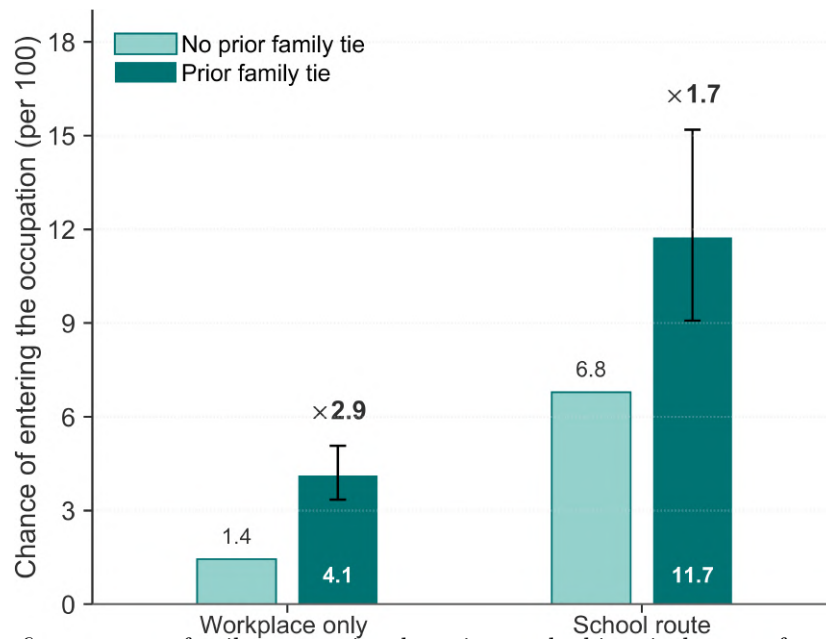
Occupation (OCC1950)	p.	Passage
<i>Workplace route only</i>		
Melters, foundry (641)	262	“As a rule, there are no apprenticeships or other organized training programs provided for melters. . . . The usual way to get one of the more skilled jobs is to begin as a furnace helper or less skilled melter and gradually to pick up the trade.”
Drop-hammer operators (524)	266	“Drop-hammer operation is learned on the job, but a regular apprenticeship is not ordinarily provided. . . . Workers usually begin as helpers and after sufficient experience advance to the job of heater.”
Inspectors, machinery parts (533)	276	“Skilled inspectors are obtained from the ranks of machine-tool operators and other machine-shop workers, or by upgrading less skilled inspectors.”
Millwrights (560)	286	“Entry into this occupation is usually through a millwright apprenticeship or equivalent on-the-job training. . . . Inexperienced workers may be hired as helpers or trainees and pick up the occupation while working.”
<i>School route also available</i>		
Painters, construction (564)	177	“Sometimes the painter learns his job by taking courses in a public trade school.”
Automobile mechanics (550)	198	“Young people under 18 should, if possible, complete at least 2 years of high school before beginning on-the-job training or entering a vocational school.”
Meat cutters (644)	355	“Vocational schools in some cities offer instruction in meat cutting.”
Carmen, railroads (553)	344	“On some roads, a few young men 18 to 25 years of age who have completed a technical high-school course are taken on as special apprentices.”

Notes: Selected passages from the 1949 *Occupational Outlook Handbook*. Page numbers refer to printed pages. Ellipses omit intervening text. The passages illustrate how workplace and school routes are coded.

through workplace-based routes.

Robustness. Table A2.3.9 reports alternative measures of predetermined family occupational experience and several robustness exercises. The difference between workplace and school-route occupations remains similar when family experience is measured by the number of prior family ties, when the sample is restricted to high-confidence occupation crosswalks, among craftsmen and when very small occupations are excluded. Reassigning occupational ties across families eliminates the relationship, providing a falsification check that the result is tied to actual family occupational histories rather than occupation-level differences alone.

Figure A2.3.4: Family occupational persistence by route of entry



Notes: The figure reports family occupational persistence by historical route of entry, estimated from equation (A2.3.1). Workplace-only occupations have no school-based entry route identified by the 1949 *Occupational Outlook Handbook*. Standard errors are two-way clustered by reconstructed family identifier and occupation. Vertical bars denote 95% confidence intervals.

Table A2.3.9: Family occupational persistence by route of entry

	β_W	β_S	$\beta_W - \beta_S$
Main specification	1.051 (0.105)	0.548 (0.131)	0.503 (0.178)
$\log(1 + \text{number of ties})$	1.468 (0.144)	0.730 (0.174)	0.738 (0.238)
Craftsmen only	1.010 (0.116)	0.537 (0.121)	0.473 (0.167)
Minimum occupation size: 1,000	1.012 (0.099)	0.549 (0.132)	0.462 (0.175)
Minimum occupation size: 5,000	0.852 (0.086)	0.541 (0.128)	0.311 (0.151)
Ties reassigned across families	-0.004 (0.009)	-0.005 (0.004)	0.001 (0.009)

Notes: PPML estimates based on equation (A2.3.1). All specifications include family-by-generation-by-year and occupation-by-year fixed effects. Standard errors are two-way clustered by reconstructed family identifier and occupation. The main specification uses an indicator for any predetermined family occupational tie. Subsequent rows vary the family-tie measure or estimation sample. The final row randomly reassigns predetermined occupational ties across families as a falsification exercise.

A2.3.4 Family ties and labor-market outcomes

The occupational-entry evidence suggests that family connections can facilitate access to information, contacts and workplace knowledge. I examine two complementary implications of this interpretation. First, employed relatives may facilitate job finding and reduce unemployment. Second, same-sector relatives may improve workers' access to sector-specific opportunities after entry, resulting in faster subsequent wage growth. I also ask whether these relationships differ between urban and non-urban labor markets, where alternative sources of information and matching may be more readily available.

For the job-finding outcomes, I estimate

$$Y_{i,t+1} = \alpha + \beta_F \text{FamTie}_{i,t} + \beta_U \text{Urban}_{i,t} + \beta_{FU} (\text{FamTie}_{i,t} \times \text{Urban}_{i,t}) + \lambda' X_{it} + \varepsilon_{i,t+1}, \quad (\text{A2.3.2})$$

where $Y_{i,t+1}$ is either an indicator for looking for work or an indicator for unemployment, $\text{FamTie}_{i,t}$ measures local employed family presence and X_{it} contains the controls and fixed effects reported in the table. The looking-for-work specification is estimated among nonworkers, while the unemployment specification uses the working-age population.

For workers, I estimate an analogous specification for subsequent wage growth,

$$\Delta \log w_{i,t+1} = \alpha + \beta_F^s \text{FamTie}_{i,t}^s + \beta_U^s \text{Urban}_{i,t} + \beta_{FU}^s (\text{FamTie}_{i,t}^s \times \text{Urban}_{i,t}) + \lambda'_s X_{it} + \varepsilon_{i,t+1}, \quad (\text{A2.3.3})$$

where $\text{FamTie}_{i,t}^s$ measures family presence in the worker's sector and the regressions are estimated separately for agricultural and manufacturing workers.

Table A2.3.10 reports the results. Among nonworkers outside urban areas, family presence is associated with a 2.87-percentage-point lower probability of looking for work. The interaction with urban status is positive, implying a smaller association of approximately 1.46 percentage points in urban areas. Family ties are also associated with a 0.60-percentage-point lower unemployment probability outside urban areas, with the association again substantially attenuated in urban labor markets. These outcomes do not measure search duration directly, but they are consistent with employed relatives facilitating job finding or reducing the incidence of unsuccessful search.

The wage results provide complementary evidence on outcomes after sector entry. Among manufacturing workers outside urban areas, a family tie is associated with approximately 2.2 log points faster subsequent hourly-wage growth. The corresponding estimate for agricultural workers is smaller and statistically indistinguishable from zero. In manufacturing, the family-tie association is substantially weaker in urban areas, as indicated by the negative interaction between urban status and family presence.

Table A2.3.10: Family ties and labor-market outcomes

VARIABLES	(1) Looking for work	(2) Unemployed Ages 15–60	(3) Δ Log hourly wage Agriculture	(4) Δ Log hourly wage Manufacturing
Urban	-0.0144*** (0.00207)	-0.00905*** (0.00144)	0.214*** (0.0603)	0.0608*** (0.00691)
Family tie	-0.02872*** (0.00145)	-0.00602*** (0.00135)	0.00916 (0.0207)	0.0222*** (0.00594)
Urban $\times$ family tie	0.0141*** (0.00210)	0.00368*** (0.00135)	-0.0443 (0.0643)	-0.0510*** (0.00742)
Observations	4,958,082	20,918,419	45,832	465,224
R-squared	0.090	0.041	0.737	0.643
Mean outcome	0.0596	0.0446		
CZ $\times$ year FE	✓	✓		
Individual controls	✓	✓	✓	✓
Cluster	CZ	CZ	CZ	CZ
Sample	Nonworkers	All	Agricultural workers	Manufacturing workers

Notes: Columns (1) and (2) relate local employed family presence to indicators for looking for work and unemployment, respectively. Column (1) is restricted to nonworkers; column (2) includes individuals ages 15–60. Columns (3) and (4) relate same-sector family presence to subsequent changes in log hourly wages among agricultural and manufacturing workers, respectively. Each specification includes urban status and its interaction with the family-tie measure. The additional individual controls and fixed effects are reported in the table. Standard errors are clustered by CZ. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A2.3.5 Persistence of sectoral differences across generations

The main text studies changes in sectoral employment following the loss of an existing sector-specific family connection. This subsection asks a complementary descriptive question: whether sectoral differences associated with an ancestor’s occupational transition remain visible among later descendants.

I begin with individuals employed in agriculture in 1860 and distinguish ancestors who entered manufacturing by 1870 from those who remained in agriculture. Let

$$T_a = \mathbf{1}\{\text{ancestor } a \text{ enters manufacturing by 1870}\},$$

and let $G_{ia} \in \{1, \dots, 5\}$ denote descendant i ’s genealogical distance from ancestor a . I estimate

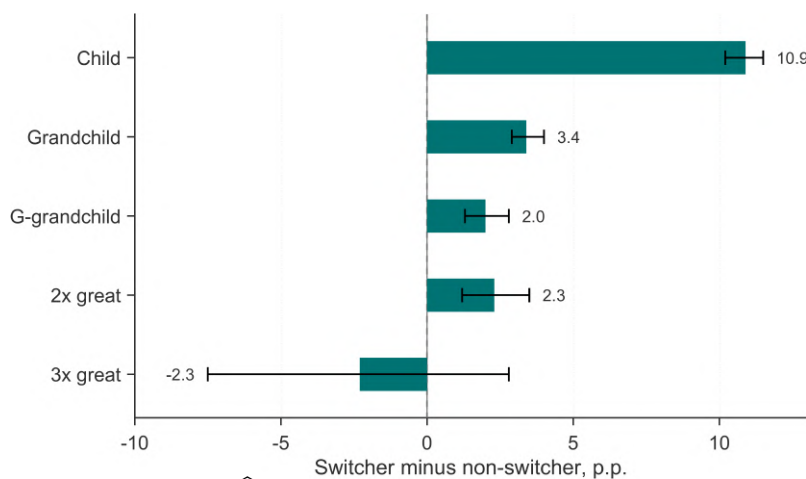
$$100 \times \mathbf{1}\{\text{sector}_{iat} = M\} = \alpha_{cat} + \lambda_{G_{ia}} + \sum_{g=1}^5 \beta_g [T_a \times \mathbf{1}\{G_{ia} = g\}] + Z_a' \Gamma_{G_{ia}} + \varepsilon_{iat}, \quad (\text{A2.3.4})$$

where α_{cat} are ancestor-origin-CZ-by-year fixed effects, $\lambda_{G_{ia}}$ are generation fixed effects and Z_a contains predetermined characteristics of ancestor a measured in 1860, including literacy, farm ownership, age, urban status and sex. The coefficients β_g therefore compare manufacturing

employment among descendants of observationally similar agricultural ancestors from the same origin CZ according to whether the ancestor subsequently entered manufacturing.

Figure A2.3.5 reports the estimates. Descendants of ancestors who entered manufacturing are approximately 9–11 percentage points more likely to work in manufacturing in the first descendant generation. The difference declines with genealogical distance but remains positive among later descendants. Sectoral differences associated with the ancestor’s transition therefore remain visible beyond the immediately following generation.

Figure A2.3.5: Manufacturing persistence across generations



Notes: The figure reports $\hat{\beta}_g$ from equation (A2.3.4), where g denotes genealogical distance from the agricultural ancestor observed in 1860. The comparison is between descendants of ancestors who entered manufacturing by 1870 and descendants of ancestors who remained in agriculture, conditional on ancestor-origin-CZ-by-year fixed effects and predetermined ancestor characteristics. Standard errors are clustered at the root-lineage level. Vertical bars denote 95% confidence intervals.

A2.4 Families and migration choice

The main text establishes two spatial margins of family connections. First, losing a non-employed relative at origin raises subsequent out-migration. Second, migration is disproportionately directed toward destinations where members of the individual's family network already live. This appendix reports additional evidence on the origin mechanism.

A2.4.1 Heterogeneity in the response to losing an origin tie

Table A2.4.1 examines heterogeneity in the out-migration response following the death of a local non-employed relative. The specifications follow equation (3) and differ only in the indicated subsample. Because the death removes a geographically proximate family connection without removing a contemporaneous employment tie, these splits help characterize the types of households for which local family support appears most valuable.

The point estimates are larger among households with young children and in rural areas. I interpret this heterogeneity cautiously: the subsamples differ along several dimensions and the estimates do not separately identify childcare, household production or informal insurance. The pattern is nevertheless consistent with locally provided family inputs being particularly valuable where household needs are greater or formal substitutes are less readily available.

Table A2.4.1: Out-migration after a family death: heterogeneity

	Dependent variable: migration						
	All (1)	W.o. 5yrs old (2)	With 5yrs old (3)	Poor (4)	Rich (5)	Urban (6)	Rural (7)
Death (non-emp.) -20 years	0.0413 (0.105)	0.0851 (0.183)	-0.216 (0.186)	-0.00706 (0.151)	0.0417 (0.149)	-0.331 (0.553)	-0.149 (0.131)
Death (non-emp.)	0.229*** (0.0724)	0.159 (0.139)	0.327*** (0.0848)	0.327*** (0.123)	0.145* (0.0862)	0.149 (0.260)	0.191** (0.0943)
Death (non-emp.) +10 years	0.426*** (0.0656)	0.375*** (0.121)	0.681*** (0.0956)	0.432*** (0.0750)	0.385*** (0.0904)	0.167 (0.282)	0.309*** (0.117)
Death (non-emp.) +20 years	0.341*** (0.0770)	0.250* (0.133)	0.546*** (0.143)	0.433*** (0.102)	0.265** (0.104)	-0.578 (0.424)	0.249 (0.161)
Observations	18,701,551	14,506,810	4,194,741	9,434,322	9,267,229	2,793,325	15,908,226
Mean outcome	19.0	19.0	19.0	19.0	19.0	19.0	19.0
Individual FE	✓	✓	✓	✓	✓	✓	✓
Individual controls	✓	✓	✓	✓	✓	✓	✓
Estimation	OLS	OLS	OLS	OLS	OLS	OLS	OLS

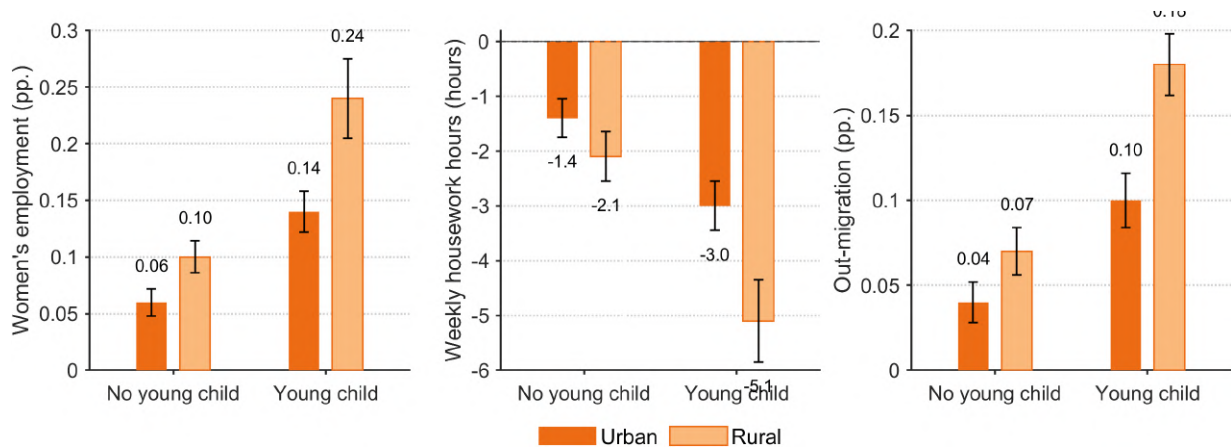
Notes: The table reports heterogeneity in the response of out-migration to the first eligible death of a local non-employed relative using the stacked event-study specification in equation (3). The outcome is an indicator for moving across commuting zones between consecutive censuses. Columns report estimates for the indicated subsamples. Event time zero is the migration interval containing the death; subsequent intervals are post-death. All specifications include individual-by-stack fixed effects and pre-event-origin-CZ-by-year-by-stack fixed effects, together with age controls and predetermined family and household characteristics. Observations are censored from the second eligible death onward. Regressions use the corrective stack weights together with the population-reweighting weights. Standard errors, reported in parentheses, are clustered at the de-duplicated family-identifier level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A2.4.2 Evidence on locally provided family support

The origin-death design is consistent with geographically proximate relatives providing services that raise the value of remaining nearby. I examine this interpretation using household outcomes for which local family support may be particularly relevant.

Figure A2.4.1 combines two pieces of evidence. First, the presence of a local non-employed relative is associated with higher female employment and fewer hours devoted to housework. These relationships are particularly pronounced among households with young children and outside urban areas. Second, the out-migration response following the death of such a relative is larger for similar households. The first set of estimates is descriptive, while the migration estimates come from the death design in equation (3).

Figure A2.4.1: Family ties, household constraints and place attachment



Notes: The figure reports relationships between local non-employed family presence and women's employment and weekly housework hours, together with the out-migration response following the death of a local non-employed relative. Estimates are shown by the presence of a child below age five and by urban status. The household-outcome estimates are descriptive associations; the migration estimates are based on equation (3). Whiskers denote 95% confidence intervals. The migration sample contains 1,119,284 observations.

Two additional exercises help distinguish geographically provided support from a general response to family bereavement. I find no comparable migration response following the death of a relative living outside the individual's CZ. The migration response is also similar among individuals who do not own their home, making a housing-asset channel less likely to account for the result. I therefore interpret the origin-death evidence as consistent with the loss of locally provided family support rather than as identifying any single underlying service.

A2.4.3 Migration and sectoral switching

The quantitative model specifies sector choice before location choice. I provide additional empirical evidence on this sequencing by asking whether migration induced by the loss of a

local non-employed relative is accompanied by movement out of the worker's initial sector.

The identifying shock is the same death of a local non-employed relative used in the origin-migration design. These deaths increase out-migration but, as shown in the sectoral placebo exercises in Appendix A2.3.2, do not generate a detectable change in sectoral employment on average. This is consistent with the shock removing locally provided family support without directly removing a sector-specific employment connection.

I focus on workers initially employed in agriculture and use the death shock as an instrument for out-migration. The first stage is

$$\text{OutMig}_{it} = \pi \text{Death}_{it} + \alpha_i + \lambda_{ct} + \gamma' X_{it} + u_{it}, \quad (\text{A2.4.1})$$

where Death_{it} captures exposure to the death of a local non-employed relative and the remaining controls follow the migration specification. I then estimate

$$\text{NonAgric}_{i,t+1} = \beta \widehat{\text{OutMig}_{it}} + \alpha_i + \lambda_{ct} + \gamma' X_{it} + \varepsilon_{it}. \quad (\text{A2.4.2})$$

Table A2.4.2 shows that the migration response generated by the family shock is not accompanied by a detectable increase in non-agricultural employment. The point estimate is small and statistically indistinguishable from zero. This result is consistent with migration operating primarily as a location-adjustment margin within a worker's existing sectoral attachment in this historical setting.

I interpret this exercise as supporting evidence for the model's sequencing rather than as a general estimate of the causal effect of migration on sector choice. The exclusion restriction requires the death of a non-employed relative to affect subsequent sector choice only through migration. The absence of a sectoral reduced-form response to these deaths supports this restriction, but does not rule out all other direct effects of losing local family support.

Table A2.4.2: Sectoral returns to migration

VARIABLES	(1) Agric.	(2) Manuf.
Migration	-0.0224 (0.0746)	0.0224 (0.0746)
Observations	2,357,596	2,357,596
R-squared	0.003	0.003
Mean outcome	0.768	0.232
Kleibergen-Paap F-statistic	16.88	16.88
First-stage R-squared	0.612	0.612
Individual FE	✓	✓
State-Year FE	✓	✓
Individual controls	✓	✓
Cluster	Individual	Individual
Pop. weighted	✓	✓
Sample	Agric./Manuf. workers	Agric./Manuf. workers

Notes: The unit of observation is an individual; standard errors are clustered at the individual level. Agric./Manuf. is an indicator for employment in agriculture or manufacturing scaled to 100 or 0. Individual controls include age, the number of family members in agriculture and manufacturing and marital status. The sample consists of children of parents working in agriculture or manufacturing and spans 1900–1940 in 10-year intervals. *** denotes significance at the 1 percent level; ** denotes significance at the 5 percent level; * denotes significance at the 10 percent level.

A3 Theory

A3.1 Stylized model

This appendix develops a two-generation model that isolates two ways in which horizontal kin connections shape adjustment to changing economic opportunities. First, inherited kin composition affects the current generation's response to changes in relative returns. I call this *inheritance*. Second, current sector and location choices reshape the kin composition inherited by the next generation. I call this *propagation*.

Environment. There are two sectors, agriculture A and manufacturing M , and two locations, an origin n and a destination m . I consider one ancestral group g over two generations. Children choose sectors at t and locations as adults at $t+1$. The example holds other continuation values fixed to isolate the feedback through horizontal kin, while sector choice accounts for the value of the subsequent location decision.³⁶

I write the model for a parental branch located at n and suppress its parental-sector index j where possible. The direct parental tie is therefore fixed when choices are made and its contribution is absorbed into branch-specific constants below. The destination-based branch is symmetric, with n and m reversed. This conditioning mirrors the quantitative model, in which sectoral kin support is evaluated at the parent's location.

Generation t inherits a predetermined horizontal-kin distribution and then faces a temporary increase in the manufacturing wage at the origin when choosing locations as adults:

$$\log w_{n,t+1}^M = \log \bar{w}_n^M + \varepsilon_t, \quad \varepsilon_{t+1} = 0.$$

The shock therefore affects the following generation only through the horizontal-kin distribution created by the current generation.

Let $\Phi_t^{\ell k} \geq 0$ denote horizontal-kin mass in location $\ell \in \{n, m\}$ and sector $k \in \{A, M\}$, with

$$\Phi_t^{nM} + \Phi_t^{nA} + \Phi_t^{mM} + \Phi_t^{mA} = r_g.$$

Define total horizontal-kin mass at each location by

$$\Phi_t^n \equiv \Phi_t^{nM} + \Phi_t^{nA}, \quad \Phi_t^m \equiv \Phi_t^{mM} + \Phi_t^{mA} = r_g - \Phi_t^n.$$

Sector choice depends on sector-specific kin at the parent's location: the origin-based branch values Φ_t^{nM} when choosing manufacturing and Φ_t^{nA} when choosing agriculture. Manufacturing

³⁶If sector and location shocks have a common EV1 scale, the same probabilities can equivalently be generated by a four-way multinomial logit over $\{nA, nM, mA, mM\}$. Different shock scales lead to the sequential formulation used here.

kin at m affect sector choice indirectly through the value of the subsequent location decision. As in the main text, write

$$P_t^{O,M} \equiv P_t^O(M \mid n, j, g), \quad P_{t+1}^{N,k} \equiv P_{t+1}^N(m \mid n, k, g).$$

The first is the probability of entering manufacturing and the second is the probability of moving to m , conditional on sector k .

Location choice. After choosing sector $k \in \{A, M\}$, an individual chooses between n and m at $t + 1$. Effective geographic kin support is

$$\begin{aligned} H_{t+1}(n) &= 1 + \psi_R^H \Phi_t^n, \\ H_{t+1}(m) &= \psi_R^H \Phi_t^m, \end{aligned}$$

where the unit term is the direct parent at the origin and $\psi_R^H > 0$ is the weight on horizontal kin relative to the parent.

Deterministic location values are

$$\begin{aligned} B_{gn,t+1}^k(n) &= \bar{B}_{gn,t+1}^k(n) + \chi_{t+1}^M \varepsilon_t \mathbf{1}\{k = M\} + \theta_H H_{t+1}(n), \\ B_{gn,t+1}^k(m) &= \bar{B}_{gn,t+1}^k(m) + \theta_H H_{t+1}(m), \end{aligned}$$

where the $\bar{B}$ -terms collect wages, migration costs and other non-kin payoffs at the normal wage, and

$$\chi_{t+1}^M \equiv \left. \frac{\partial v(w_{n,t+1}^M, p_{n,t+1})}{\partial \log w_{n,t+1}^M} \right|_{\varepsilon_t=0} > 0.$$

The expression for the wage response is evaluated locally around $\varepsilon_t = 0$.

With Type-I extreme-value location shocks of scale σ_N ,

$$P_{t+1}^{N,k} = \frac{\exp(B_{gn,t+1}^k(m)/\sigma_N)}{\exp(B_{gn,t+1}^k(n)/\sigma_N) + \exp(B_{gn,t+1}^k(m)/\sigma_N)}.$$

Let $a_{t+1}^{N,k}$ collect the fixed relative value of the destination, including the direct-parent contribution. Since $\Phi_t^n = r_g - \Phi_t^m$, the destination-versus-origin log odds are

$$\log \frac{P_{t+1}^{N,k}}{1 - P_{t+1}^{N,k}} = a_{t+1}^{N,k} - \frac{\chi_{t+1}^M}{\sigma_N} \varepsilon_t \mathbf{1}\{k = M\} + \frac{\theta_H \psi_R^H}{\sigma_N} (2\Phi_t^m - r_g).$$

Holding total horizontal-kin mass fixed,

$$\frac{\partial P_{t+1}^{N,k}}{\partial \Phi_t^m} = \frac{2\theta_H \psi_R^H}{\sigma_N} P_{t+1}^{N,k} (1 - P_{t+1}^{N,k}) > 0.$$

For manufacturing,

$$\frac{\partial P_{t+1}^{N,M}}{\partial \varepsilon_t} = -\frac{\chi_{t+1}^M}{\sigma_N} P_{t+1}^{N,M} (1 - P_{t+1}^{N,M}) < 0,$$

while $\partial P_{t+1}^{N,A}/\partial \varepsilon_t = 0$. A temporary increase in the origin manufacturing wage therefore raises the incentive of manufacturing workers to remain at n .

Sector choice. Sector choice accounts for the subsequent location decision. The location inclusive value is

$$V_{t+1}(n, k, g) = \sigma_N \log [\exp(B_{gn,t+1}^k/\sigma_N) + \exp(B_{gn,t+1}^m/\sigma_N)].$$

Consistent with the quantitative model, effective sector-specific kin support at the parent's location is

$$F_t^M = \mathbf{1}\{j = M\} + \psi_R^M \Phi_t^{nM}, \quad F_t^A = \mathbf{1}\{j = A\} + \psi_R^A \Phi_t^{nA},$$

where $\psi_R^k > 0$ is the weight on horizontal sector- k kin relative to the parent. Sector values are

$$\mathcal{U}_{gnjt}^k = \bar{\mathcal{U}}_{gnjt}^k + \theta_k F_t^k + V_{t+1}(n, k, g).$$

With Type-I extreme-value sector shocks with dispersion σ_O ,

$$P_t^{O,M} = \frac{\exp(\mathcal{U}_{gnjt}^M/\sigma_O)}{\exp(\mathcal{U}_{gnjt}^M/\sigma_O) + \exp(\mathcal{U}_{gnjt}^A/\sigma_O)}.$$

After absorbing the direct parental-sector contribution into a_t^O , manufacturing entry satisfies

$$\log \frac{P_t^{O,M}}{1 - P_t^{O,M}} = a_t^O + \frac{\theta_M \psi_R^M}{\sigma_O} \Phi_t^{nM} - \frac{\theta_A \psi_R^A}{\sigma_O} \Phi_t^{nA},$$

where a_t^O includes the difference in location inclusive values.

Let $s_t^M \equiv \Phi_t^{nM}/\Phi_t^n$ denote the manufacturing share of local horizontal kin, for $\Phi_t^n > 0$. Holding the geographic distribution of kin fixed, so that Φ_t^n is constant,

$$\frac{\partial P_t^{O,M}}{\partial s_t^M} = \frac{\Phi_t^n (\theta_M \psi_R^M + \theta_A \psi_R^A)}{\sigma_O} P_t^{O,M} (1 - P_t^{O,M}) > 0.$$

Kin geography also affects sector choice through the location inclusive values. Along a change

in destination kin mass, holding the local sector-specific contributions fixed,

$$\left. \frac{\partial P_t^{O,M}}{\partial \Phi_t^m} \right|_{\text{location channel}} = \frac{2\theta_H \psi_R^H}{\sigma_O} \left(P_{t+1}^{N,M} - P_{t+1}^{N,A} \right) P_t^{O,M} \left(1 - P_t^{O,M} \right).$$

Its sign depends on which sector is more destination-oriented.

Since

$$\frac{\partial V_{t+1}(n, M, g)}{\partial \varepsilon_t} = \chi_{t+1}^M \left(1 - P_{t+1}^{N,M} \right),$$

the contemporaneous manufacturing response is

$$\frac{\partial P_t^{O,M}}{\partial \varepsilon_t} = \frac{\chi_{t+1}^M}{\sigma_O} \left(1 - P_{t+1}^{N,M} \right) P_t^{O,M} \left(1 - P_t^{O,M} \right) > 0.$$

Evolution of kin composition. With a continuum of descendants, current choice probabilities equal cohort shares. Those shares determine the horizontal-kin distribution faced by the next generation:

$$\begin{aligned} \Phi_{t+1}^{nM} &= r_g P_t^{O,M} \left(1 - P_{t+1}^{N,M} \right), & \Phi_{t+1}^{mM} &= r_g P_t^{O,M} P_{t+1}^{N,M}, \\ \Phi_{t+1}^{nA} &= r_g (1 - P_t^{O,M}) \left(1 - P_{t+1}^{N,A} \right), & \Phi_{t+1}^{mA} &= r_g (1 - P_t^{O,M}) P_{t+1}^{N,A}. \end{aligned}$$

These masses sum to r_g . Aggregating across sectors gives

$$\Phi_{t+1}^m = r_g \left[P_t^{O,M} P_{t+1}^{N,M} + (1 - P_t^{O,M}) P_{t+1}^{N,A} \right]. \quad (\text{A3.1.1})$$

For an origin-based descendant, the sectoral kin inputs in the next generation are Φ_{t+1}^{nM} and Φ_{t+1}^{nA} . For a destination-based descendant, the corresponding local inputs are Φ_{t+1}^{mM} and Φ_{t+1}^{mA} . The quantitative model keeps both branches and averages them using the appropriate parental-state distribution; the stylized derivations below report the origin-based branch to keep the mechanism transparent.

Inheritance. Inheritance captures how the predetermined kin composition changes the current response to the wage shock. Consider a marginal shock around $\varepsilon_t = 0$ and define

$$\begin{aligned} R_t^O &\equiv \left. \frac{\partial P_t^{O,M}}{\partial \varepsilon_t} \right|_{\varepsilon_t=0} = \frac{\chi_{t+1}^M}{\sigma_O} \left(1 - P_{t+1}^{N,M} \right) P_t^{O,M} \left(1 - P_t^{O,M} \right), \\ R_t^N &\equiv - \left. \frac{\partial P_{t+1}^{N,M}}{\partial \varepsilon_t} \right|_{\varepsilon_t=0} = \frac{\chi_{t+1}^M}{\sigma_N} P_{t+1}^{N,M} \left(1 - P_{t+1}^{N,M} \right). \end{aligned} \quad (\text{A3.1.2})$$

Proposition 1: inherited kin composition shapes responsiveness.

Holding kin geography fixed when varying local sector composition,

$$\begin{aligned}\frac{\partial R_t^O}{\partial s_t^M} &= \frac{\chi_{t+1}^M}{\sigma_O} \left(1 - P_{t+1}^{N,M}\right) \frac{\Phi_t^n (\theta_M \psi_R^M + \theta_A \psi_R^A)}{\sigma_O} P_t^{O,M} \left(1 - P_t^{O,M}\right) \left(1 - 2P_t^{O,M}\right), \\ \frac{\partial R_t^N}{\partial \Phi_t^m} &= \frac{\chi_{t+1}^M}{\sigma_N} \frac{2\theta_H \psi_R^H}{\sigma_N} P_{t+1}^{N,M} \left(1 - P_{t+1}^{N,M}\right) \left(1 - 2P_{t+1}^{N,M}\right).\end{aligned}\quad (\text{A3.1.3})$$

Hence, when $P_t^{O,M} < 1/2$, replacing a local agricultural connection with a local manufacturing connection increases the manufacturing response. When $P_{t+1}^{N,M} < 1/2$, shifting horizontal kin from the origin toward the destination increases the magnitude of the migration response.

Proof. Holding kin geography fixed,

$$\frac{\partial R_t^O}{\partial s_t^M} = \frac{\chi_{t+1}^M}{\sigma_O} \left(1 - P_{t+1}^{N,M}\right) \left(1 - 2P_t^{O,M}\right) \frac{\partial P_t^{O,M}}{\partial s_t^M},$$

which gives the first result. Similarly,

$$\frac{\partial R_t^N}{\partial \Phi_t^m} = \frac{\chi_{t+1}^M}{\sigma_N} \left(1 - 2P_{t+1}^{N,M}\right) \frac{\partial P_{t+1}^{N,M}}{\partial \Phi_t^m},$$

which gives the second. $\square$

The result follows from logit curvature. Local sector composition shifts the baseline manufacturing probability, while kin geography shifts the baseline destination probability. When a choice is initially uncommon, connections pointing toward that choice raise responsiveness; once its baseline probability exceeds one half, the sign reverses.

Propagation. Propagation arises because the current generation's choices change the joint sector-location kin distribution inherited by the next generation. Differentiating the laws above,

$$\begin{aligned}G_t^{nM} &\equiv \frac{\partial \Phi_{t+1}^{nM}}{\partial \varepsilon_t} = r_g \left[\left(1 - P_{t+1}^{N,M}\right) R_t^O + P_t^{O,M} R_t^N \right] > 0, \\ G_t^{mA} &\equiv \frac{\partial \Phi_{t+1}^{nA}}{\partial \varepsilon_t} = -r_g \left(1 - P_{t+1}^{N,A}\right) R_t^O < 0, \\ G_t^m &\equiv \frac{\partial \Phi_{t+1}^m}{\partial \varepsilon_t} = r_g \left[\left(P_{t+1}^{N,M} - P_{t+1}^{N,A}\right) R_t^O - P_t^{O,M} R_t^N \right].\end{aligned}\quad (\text{A3.1.4})$$

The shock therefore creates more local manufacturing kin and fewer local agricultural kin for the next generation's origin-based branch. Its effect on total destination kin is ambiguous: the first term in G_t^m reflects the change in sector composition and the second the decline in destination choice among manufacturing workers.

Because the wage shock has disappeared when the next generation makes its location choices,

that generation responds only through its inherited kin distribution:

$$\begin{aligned} \frac{\partial P_{t+1}^{O,M}}{\partial \varepsilon_t} = & P_{t+1}^{O,M} \left(1 - P_{t+1}^{O,M} \right) \left[\frac{\theta_M \psi_R^M}{\sigma_O} G_t^{mM} - \frac{\theta_A \psi_R^A}{\sigma_O} G_t^{mA} \right. \\ & \left. + \frac{2\theta_H \psi_R^H}{\sigma_O} \left(P_{t+2}^{N,M} - P_{t+2}^{N,A} \right) G_t^m \right]. \end{aligned} \quad (\text{A3.1.5})$$

The first two terms are local sectoral propagation and are positive after a positive origin-manufacturing wage shock. The third is spatial propagation and can reinforce or offset them.

Proposition 2: kin connections can amplify or dampen cumulative adjustment.

Define the cumulative two-generation manufacturing response with horizontal kin as

$$\mathcal{C}_F^M \equiv \frac{\partial \left(P_t^{O,M} + P_{t+1}^{O,M} \right)}{\partial \varepsilon_t}.$$

Consider a matched no-horizontal-kin benchmark in which the horizontal-kin utility coefficients are set to zero and branch-specific constants are adjusted so that the no-shock probabilities $P_t^{O,M}$, $P_{t+1}^{N,M}$ and $P_{t+1}^{N,A}$ match those in the economy with horizontal kin. The impact response is then identical across economies, while the benchmark has no horizontal-kin propagation: $\mathcal{C}_0^M = R_t^O$.

For the origin-based branch, the difference in cumulative adjustment is

$$\begin{aligned} \mathcal{C}_F^M - \mathcal{C}_0^M = & P_{t+1}^{O,M} \left(1 - P_{t+1}^{O,M} \right) \left[\frac{\theta_M \psi_R^M}{\sigma_O} G_t^{mM} - \frac{\theta_A \psi_R^A}{\sigma_O} G_t^{mA} \right. \\ & \left. + \frac{2\theta_H \psi_R^H}{\sigma_O} \left(P_{t+2}^{N,M} - P_{t+2}^{N,A} \right) G_t^m \right]. \end{aligned} \quad (\text{A3.1.6})$$

Horizontal kin therefore amplify cumulative manufacturing adjustment if

$$\frac{\theta_M \psi_R^M}{\sigma_O} G_t^{mM} - \frac{\theta_A \psi_R^A}{\sigma_O} G_t^{mA} + \frac{2\theta_H \psi_R^H}{\sigma_O} \left(P_{t+2}^{N,M} - P_{t+2}^{N,A} \right) G_t^m > 0,$$

and dampen it if the expression is negative.

Proof. In the matched no-horizontal-kin economy, the impact response is R_t^O and $\partial P_{t+1}^{O,M,0} / \partial \varepsilon_t = 0$. Hence

$$\mathcal{C}_F^M - \mathcal{C}_0^M = \frac{\partial P_{t+1}^{O,M}}{\partial \varepsilon_t}.$$

Substituting equation (A3.1.5) gives equation (A3.1.6). Since $P_{t+1}^{O,M}(1 - P_{t+1}^{O,M}) > 0$, the sign is determined by the expression in brackets. $\square$

Proposition 2 separates two propagation channels. The term

$$\frac{\theta_M \psi_R^M}{\sigma_O} G_t^{nM} - \frac{\theta_A \psi_R^A}{\sigma_O} G_t^{mA} > 0$$

captures local sectoral propagation: the shock creates more manufacturing kin and fewer agricultural kin at the origin. The term

$$\frac{2\theta_H \psi_R^H}{\sigma_O} \left(P_{t+2}^{N,M} - P_{t+2}^{N,A} \right) G_t^m$$

captures spatial propagation: the shock also changes where kin are located and this affects sector choice through future location opportunities. Spatial propagation reinforces the local sectoral channel when the induced geographic change favors the more destination-oriented sector and offsets it otherwise.

The matched benchmark isolates propagation by construction. Without adjusting the initial allocation to match, the cumulative difference also contains inheritance:

$$\mathcal{C}_F^M - \mathcal{C}_0^M = \left(R_t^{O,F} - R_t^{O,0} \right) + \frac{\partial P_{t+1}^{O,M,F}}{\partial \varepsilon_t}.$$

Proposition 1 characterizes the first term and Proposition 2 the second. Kin can therefore dampen the initial response while amplifying cumulative adjustment, or the reverse, depending on inherited local sectoral composition and the direction of subsequent propagation.

A3.2 Model extensions

The baseline deliberately restricts kin to the two margins identified most directly in the empirical analysis: local sector-specific kin lower sector-entry barriers through Q_{gnjt}^k , while geographically proximate kin raise location utility through $H_{gnjt}(m)$. This appendix considers three extensions. The first allows sector-specific kin to transmit information about earnings. The second allows kin-provided home support to substitute for market-provided household services and to affect effective labor supply. The third introduces schooling as a formal substitute for inherited sector-specific connections. In each case, the extension is written to nest the corresponding baseline specification.

A3.2.1 Kin wage learning

The baseline lets kin affect sector entry through access to information, referrals and sector-specific know-how, but market wages themselves contain no kin component. A related possibility is that relatives reveal information about the earnings opportunities available in a sector.

This extension adds such a wage-learning channel while leaving firm-side market wages and the production block unchanged.

Consider a child of ancestral type g whose parent is in location n and sector j . For sector k , define the wage signal inherited from the parent and the parent's horizontal kin as

$$W_{gnjt}^{L,k} = \mathbf{1}\{j = k\} w_{nt}^k + \psi_R^{W,k} \sum_{m \in N} \mathcal{K}_{nm}^W \Phi_{gnjt}^k(m) w_{mt}^k, \quad (\text{A3.2.1})$$

where

$$\mathcal{K}_{nm}^W = \exp(-\kappa_W \text{dist}_{nm}), \quad \kappa_W \geq 0, \quad (\text{A3.2.2})$$

is a distance-decay kernel. The parental contribution is normalized to one, consistently with the baseline sector-access stock. The parameter $\psi_R^{W,k}$ measures the strength of one unit of expected horizontal kin exposure relative to the parental wage signal.

Equation (A3.2.1) differs deliberately from the baseline sector-access stock F_{gnjt}^k . Direct sector access in equation (8) is local: horizontal kin must be in the parent's region and sector k . Wage information can instead travel across space, although the signal becomes weaker with distance. Thus, a relative working in sector k elsewhere need not provide a local job connection to transmit information about sector- k earnings.

I let this information raise the perceived earnings value of sector k . The sector-choice value in equation (7) becomes

$$\mathcal{U}_{gnjt}^{k,L} = Q_{gnjt}^k - \tau_{gnjt}^k + V_{t+1}(n, k, g) + \gamma_k^W W_{gnjt}^{L,k}, \quad \gamma_k^W \geq 0. \quad (\text{A3.2.3})$$

The term $\gamma_k^W W_{gnjt}^{L,k}$ is an expected-earnings shifter rather than an individual-specific market wage. Market wages w_{mt}^k , production and goods-market clearing therefore remain as in the baseline model. Setting $\gamma_A^W = \gamma_M^W = 0$ recovers the baseline sector-choice problem exactly.

The empirical wage evidence can be used to discipline this extension. The association between parental sector exposure and subsequent earnings progression disciplines γ_k^W , while variation in the wages and locations of horizontal kin disciplines $\psi_R^{W,k}$ and the spatial decay κ_W . In particular, wage exposure generated by relatives outside the individual's own region helps distinguish transmission through kin from contemporaneous local wage conditions. I interpret this evidence as supporting the wage-learning channel.

A3.2.2 Market substitutes for kin-provided home support

In the baseline, geographically proximate kin enter location utility directly through $\theta_H H_{gnjt}(m)$. This reduced-form amenity can capture childcare, household production, informal insurance, transportation and other forms of local support. I now allow these services to be provided either by kin or by the market and allow cheaper home support to free household time for

market work.

Let the shadow price of market-provided home support at destination m be

$$p_{mt}^h = \bar{w}_{m,t-1}^{-\xi_h}, \quad \xi_h > 0, \quad (\text{A3.2.4})$$

where $\bar{w}_{m,t-1}$ is the lagged employment-weighted average wage. Market substitutes are therefore cheaper in more developed local economies.

To preserve the linear kin amenity in the baseline model, define the shadow price of family-provided support as

$$p_{gnjt}^F(m) = \exp(-H_{gnjt}(m)). \quad (\text{A3.2.5})$$

A larger value of $H_{gnjt}(m)$ therefore represents greater access to family-provided support and a lower implicit price of obtaining that support.

Family and market support are imperfect substitutes. Their CES unit-cost index is

$$P_{gnjt}^H(m) = [\omega_f^H (p_{gnjt}^F(m))^{1-\sigma_H} + \omega_m^H (p_{mt}^h)^{1-\sigma_H}]^{\frac{1}{1-\sigma_H}}, \quad (\text{A3.2.6})$$

where $\omega_f^H + \omega_m^H = 1$ and $\sigma_H > 1$.

The corresponding family share is

$$s_{gnjt}^H(m) = \omega_f^H \left(\frac{p_{gnjt}^F(m)}{P_{gnjt}^H(m)} \right)^{1-\sigma_H}. \quad (\text{A3.2.7})$$

Cheaper home support can additionally free household time for market work. Let effective labor supplied by a worker choosing destination m be

$$e_{gnjt}^H(m) = 1 + \kappa_E (P_{gnjt}^H(m))^{-1}, \quad \kappa_E \geq 0, \quad (\text{A3.2.8})$$

so that realized labor income is

$$\tilde{w}_{gnjt}^j(m) = w_{mt}^j e_{gnjt}^H(m). \quad (\text{A3.2.9})$$

The location value in equation (11) is then replaced by

$$\begin{aligned} B_{gnt}^{j,H}(m) = & \mathcal{A}_{gmt}^j - \rho \log L_{mt} + v(\tilde{w}_{gnjt}^j(m), p_{mt}) \\ & - \theta_H \log P_{gnjt}^H(m) - M_{gnmt}^j + \beta O_t(m, j, g). \end{aligned} \quad (\text{A3.2.10})$$

This specification nests the baseline exactly. If $\kappa_E = 0$ and $\omega_m^H = 0$, then $\omega_f^H = 1$,

$$P_{gnjt}^H(m) = p_{gnjt}^F(m) = \exp(-H_{gnjt}(m)),$$

and therefore

$$-\theta_H \log P_{gnjt}^H(m) = \theta_H H_{gnjt}(m),$$

while $\tilde{w}_{gnjt}^j(m) = w_{mt}^j$. Equation (A3.2.10) consequently collapses to equation (11).

When $\omega_m^H > 0$, the market provides a substitute for proximity to kin. Conditional on family exposure, the family share $s_{gnjt}^H(m)$ is larger where market support is expensive, which in this specification occurs in lower-wage local economies. The extension therefore formalizes the idea that family proximity is especially valuable where formal substitutes for childcare and household support are scarce (Compton & Pollak, 2014).

If the effective-labor-supply margin is implemented in general equilibrium, the same factor $e_{gnjt}^H(m)$ also scales the labor units supplied to firms, so that household earnings and effective labor input move consistently. Setting $\kappa_E = 0$ shuts down this additional labor-supply channel while retaining substitution between family and market home support.

A3.2.3 Schooling as a formal substitute for kin

The baseline sector-entry value is linear in the inherited family stock, $Q_{gnjt}^k = \theta_k F_{gnjt}^k$. I now introduce schooling as a formal alternative to this inherited route. The extension is constructed so that shutting down the formal route returns the baseline family contribution exactly.

Let $\mathcal{E}_{nt} > 0$ denote the stock of formal school resources available to a child whose parent lives in n . Because $F_{gnjt}^k = \log(\exp(F_{gnjt}^k))$, I represent the family route by the positive input $\mathcal{F}_{gnjt}^k \equiv \exp(F_{gnjt}^k)$. Family and formal inputs are combined through

$$I_{gnjt}^k = [\omega_F^k \exp(\chi F_{gnjt}^k) + (1 - \omega_F^k) \mathcal{E}_{nt}^\chi]^{1/\chi}, \quad \chi \equiv \frac{\eta - 1}{\eta}, \quad (\text{A3.2.11})$$

where $\eta > 1$ is the elasticity of substitution between inherited family access and formal school resources. The sector-entry contribution becomes

$$Q_{gnjt}^{k,S} = \theta_k \log I_{gnjt}^k. \quad (\text{A3.2.12})$$

This normalization preserves the baseline. As $\omega_F^k \rightarrow 1$, equation (A3.2.11) gives $I_{gnjt}^k = \exp(F_{gnjt}^k)$, and therefore $Q_{gnjt}^{k,S} = \theta_k F_{gnjt}^k = Q_{gnjt}^k$. When $\eta > 1$, schooling and inherited family access are gross substitutes: an expansion of formal resources lowers the marginal value of an additional sector-specific family connection.

I allow school resources to increase with local development while retaining residual variation in local school supply:

$$\mathcal{E}_{nt} = \mathcal{S}_{nt} \bar{w}_{n,t-1}^{\xi_s}, \quad \xi_s > 0, \quad (\text{A3.2.13})$$

where $\bar{w}_{n,t-1}$ is the lagged employment-weighted average wage and $\mathcal{S}_{nt}$ captures school-supply variation not mechanically generated by local wages. Using the lagged wage prevents the child's

contemporaneous sector choice from feeding directly back into the formal input entering that choice.

The extension changes only the sector-entry contribution Q_{gnjt}^k . The kin state $\Phi_{g,t}$, the kin law of motion, the adult location problem, production and trade remain unchanged. Schooling nevertheless affects future kin geography indirectly because changes in sector choices alter the distribution of descendants that enters the law of motion for $\lambda_{g,t+1}$.

Estimation. The extension adds ξ_s , the CES curvature χ , the sector-specific family weights ω_F^k and the corresponding sector-entry value scales. I estimate ξ_s from the relationship between local wage growth and school expenditure per enrolled child, instrumenting wage growth with baseline industry shares interacted with leave-one-CZ-out national industry wage growth. The preferred specification yields $\hat{\xi}_s = 0.337$, with a first-stage F -statistic of 10.863.

To identify substitution, define $q_i^k = F_i^k$, $e_i = \log \mathcal{E}_i$ and $x_i^k = q_i^k - e_i$. The CES elasticity of the sector-entry index with respect to the family input is

$$s_{Fi}^k = \text{invlogit}(\alpha_k + \chi x_i^k), \quad \alpha_k = \log \frac{\omega_F^k}{1 - \omega_F^k}. \quad (\text{A3.2.14})$$

Deaths of same-sector relatives identify the family margin at different values of x_i^k . I estimate the death-induced family response in twelve pre-determined relative-effectiveness groups and fit

$$\hat{m}_{F,r}^k = \Lambda_k \text{invlogit}(\alpha_k + \chi \bar{x}_r^k) + \varepsilon_r^k. \quad (\text{A3.2.15})$$

Variation in the family-loss response across relative effectiveness identifies χ ; the horizontal location of the schedule identifies ω_F^k ; and its level identifies Λ_k . I impose a common curvature across sectors.

School expansion and its interaction with family geography I use the extension to repeat the dynamic family accounting for the historical increase in school expenditure. The exercise compares the response without family channels, with inherited horizontal family geography held fixed and in the full model with endogenous family renewal.

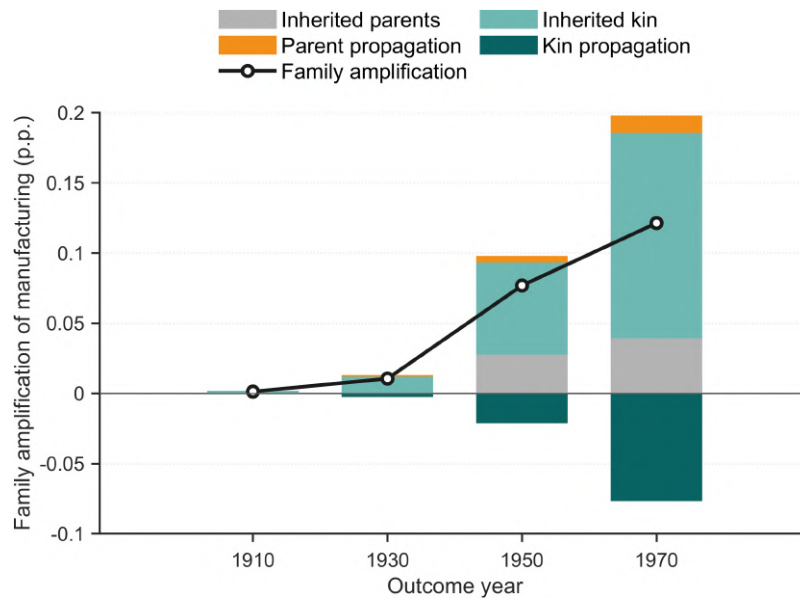
In 1950, higher school expenditure raises the manufacturing share by 0.122 percentage points without families and by 0.199 points in the full model, a 63 percent larger response. Inherited parental and horizontal ties contribute 0.028 and 0.066 points, while propagation subtracts 0.016 points. By 1970, the response is 0.170 points without families and 0.291 points in the full model. Inherited incidence contributes about 0.186 points, while subsequent renewal subtracts 0.064 points. Schooling therefore draws primarily on sectoral family connections already in place.

Table A3.2.1: Sector-entry parameters

	Agriculture	Manufacturing
<i>Panel A. CES curvature</i>		
Curvature, $\hat{\chi}_k$	0.857** (0.289)	0.936** (0.274)
Number of groups	12	12
First-stage observations	318,287	74,162
<i>Panel B. Common curvature and elasticity</i>		
Common curvature, $\hat{\chi}$		0.899*** (0.199)
Implied substitution elasticity, $\hat{\eta} = 1/(1 - \hat{\chi})$		9.86
<i>Panel C. Sector-specific entry parameters</i>		
Scale-free sector-entry value, $\hat{\Lambda}_k$	0.0698	0.0724
CES family weight, $\hat{\omega}_F^k$	0.584	0.385

Notes: Panel A reports the unrestricted sector-specific curvature estimates $\hat{\chi}_k$ from equation (A3.2.15). Panel B reports the common curvature $\hat{\chi}$. Panel C reports $\hat{\Lambda}_k$ and $\hat{\omega}_F^k$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A3.2.1: Schooling and kin capital



Notes: Bars allocate the family contribution between inherited parental and horizontal ties and endogenous parental and horizontal propagation.

A4 Quantification

This section describes the data sources, the construction of the variables used in the empirical analysis and the procedures undertaken to estimate the structural parameters of the model for

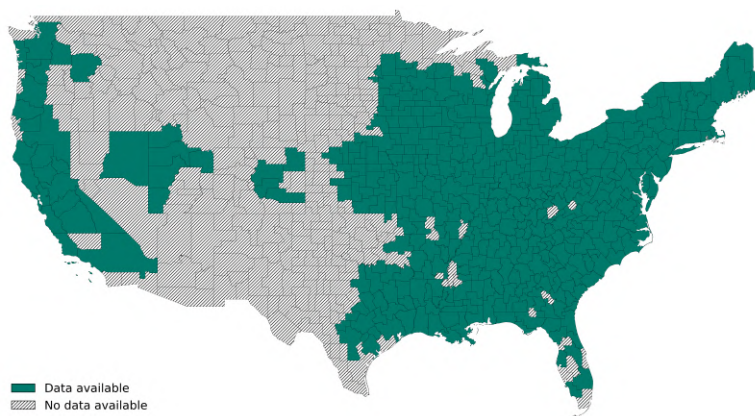
the U.S. between 1880 and 1950.

A4.1 Data

The primary data source is the complete-count U.S. Population Census from 1850 to 1950, accessed through IPUMS USA (Ruggles et al., 2024) in order to construct migration flows and parental-child industry transitions. In addition, information on agricultural wages is obtained from the Agricultural Census (1870–1940) and information on manufacturing salaries from the Manufacturing Census (1880–1940), both accessed through IPUMS NHGIS (Schroeder et al., 2025). To harmonize the geographic structure over time, consistent county-to-commuting-zone crosswalks developed by Eckert et al. (2020) are employed.

Regions. I follow the regional definition of Eckert and Peters (2022) and use commuting zones (CZs). County-to-CZ crosswalks are from Eckert et al. (2020) and respect historical county boundaries. A balanced sample from 1880 to 1950 yields 455 CZs. Table A4.1.1 compares the model sample with the full U.S. economy. The agricultural share among agricultural and manufacturing workers aged 20–40 is close to the national share early in the sample, although the fit deteriorates over time. The model sample contains roughly 23–25% of the U.S. population.

Figure A4.1.1: Region sample



Notes: This figure shows the 455 commuting zones that are used in the quantitative analysis. Source: Eckert et al. (2020).

Sectoral and age definition. I focus on agriculture and manufacturing, the two sectors that account for most employment over the historical period and for which comparable wage measures are available (Eckert & Peters, 2022). Workers are classified using industry (IND1950) codes. Agricultural workers have industry code 105. Manufacturing workers have industry

Table A4.1.1: Sectoral employment and population in model vs. census, 1880–1950

	Agric. share	Employment share		Population
Year	among A+M %	Agric. %	Manuf. %	millions
<i>Panel A. Model commuting zones, ages 20–40</i>				
1880	68.3	36.8	17.1	12.54
1910	43.8	19.8	25.4	23.20
1930	28.6	11.7	29.3	29.48
1950	18.2	7.4	33.4	35.07
<i>Panel B. United States, ages 20–40</i>				
1880	75.8	44.2	14.1	16.29
1910	58.2	28.8	20.7	32.25
1930	42.9	18.6	24.8	41.13
1950	28.6	11.4	28.5	49.54
<i>Panel C. United States, all ages</i>				
1880	78.6	49.0	13.3	49.76
1910	65.2	35.3	18.8	92.02
1930	50.9	23.5	22.7	122.76
1950	35.6	14.3	25.9	152.26

Notes: The table reports sectoral employment shares and population totals for the model sample and the full United States. The first column reports agricultural employment as a share of agricultural plus manufacturing employment. The next two columns report agricultural and manufacturing employment as shares of the relevant population. Population is reported in millions.

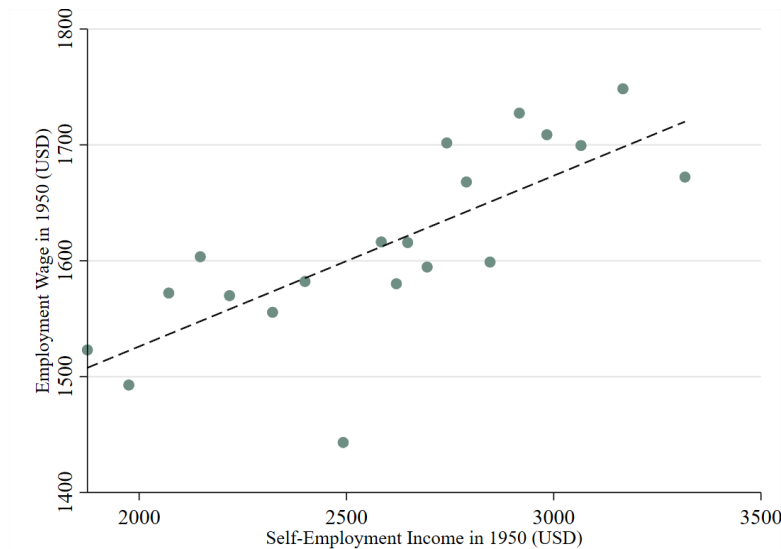
codes 306–499, covering durable and non-durable manufacturing. Construction and mining are excluded because wages for these groups are not consistently observed before 1940. Children are individuals aged 0–20. Parents are individuals aged 20–40 with at least one co-resident child (`NCHILD > 0`).

Migration Flows. Migration flows are measured using individual-level links from the Census Linking Tree Project ([Buckles et al., 2025](#)) for 1880–1910, 1910–1930 from IPUMS Multigenerational Longitudinal Project for 1930–1950 ([Helgertz et al., 2023](#)). Parents are defined as individuals aged 20–40 with at least one co-resident child (`NCHILD > 0`). Gross migration flows and migration probabilities are constructed by following individuals 20/30 years in the future with a single crosswalk (no chained crosswalks over two consecutive 10-year periods).³⁷ Migration flows across regions are computed by aggregating them to the CZ level.

³⁷This is done for consistency because there is no census data for 1890.

Wages. Manufacturing wages for 1880–1920 are obtained from the Manufacturing Census (NHGIS) as total wages divided by the number of workers, aggregated to the commuting-zone level. Since manufacturing wages are not available for 1930, I use the average wages from the 1920 Manufacturing Census and the 1940 individual census. For 1950, individual-level census data allow direct computation of average wages for manufacturing workers aged 20–40. Agricultural wages are measured as expenditures on farm labor (1870, 1900, 1920) per county, divided by the number of farm laborers aged 20–60. The model covers slightly different time periods so I use the wages in 1870 for the 1880 period, the wages in 1900 for 1910 period and the wages in 1920 for 1930. While self-employed farmers’ incomes are excluded, validation against Iowa state census data and comparisons with 1950 commuting-zone income distributions suggest the measure is reasonable. In cases where individual wages are present, I do not residualize them using Mincer regressions so that they are consistent with the averages for 1880–1920.

Figure A4.1.2: Agricultural income comparison: self-employed (total) vs wage income

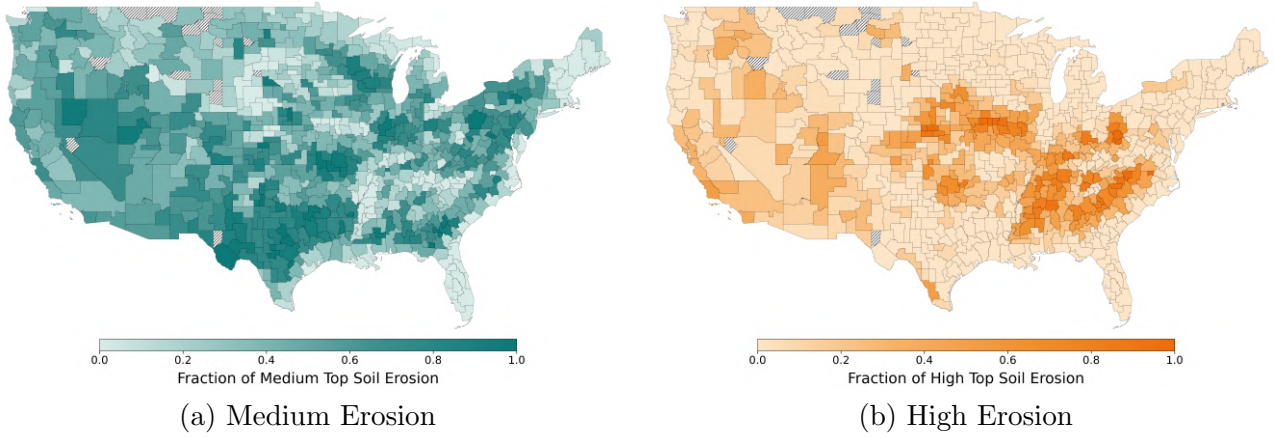


Notes: This figure shows a binned scatter plot between self-employed income from farmer individuals and wage income from the same industry group within CZs in 1950. The dotted line represents the fitted line. Observations: 227 CZs. Source: [Eckert et al. \(2020\)](#); [Helgertz et al. \(2023\)](#).

Productivity. Figure [A4.1.3](#) depicts the geographical distribution of soil erosion taken from [Hornbeck \(2012\)](#).

Trade Costs. I construct bilateral trade costs between regions using the county-to-county transportation cost data from [Hornbeck and Rotemberg \(2024\)](#). Their data digitize the U.S. transportation network by decade from 1830 to 1920 and report the cost of shipping one ton of goods between county centroids, holding counties fixed at 1890 boundaries. Following [Hornbeck](#)

Figure A4.1.3: Geographical Distribution of Soil Erosion



Notes: Panel (a) shows the geographical distribution of medium soil erosion across CZs per agricultural area. Panel (b) shows the geographical distribution of high soil erosion across CZs per agricultural area. High degrees of soil erosion are defined as regions with more than 75 percent of topsoil lost and potential numerous or deep gullies, while medium erosion is characterized by areas with 25 percent to 75 percent of topsoil lost and potentially some gullies (Hornbeck, 2012).

and Rotemberg (2024), I convert dollar-per-ton costs into iceberg trade costs by assuming an average value of \$35 per ton of agricultural goods in 1890 following Eckert and Peters (2022):

$$T_{odt} = 1 + \frac{t_{odt}}{35}, \quad (\text{A4.1.1})$$

where t_{odt} is the reported transportation cost from county o to county d in decade t . I then aggregate the resulting county-level trade cost matrix to the regional definition.

A4.2 Construction of migration distance

Migration costs in the model are governed by an elasticity on the distance between CZ pairs. Rather than using straight-line distance between CZ centroids, I construct a rail-adjusted effective distance that accounts for the historical railroad network, following the least-cost-path approach used in the market-access literature (Donaldson, 2018; Donaldson & Hornbeck, 2016). This approach is motivated by evidence that rail access shaped population movement and settlement directly, not only goods trade (Atack, Bateman, Haines, & Margo, 2010).

Railroad segments are taken from the Atack (2016) historical railroad shapefile, which records the year each segment entered operation (1826–1916) along with its geographic path. Commuting zones are the 1990 boundaries used throughout the paper (Eckert et al., 2020; Tolbert & Sizer, 1996), for which I take the polygon centroid as the CZ’s location.

For each migration wave $t \in \{1880, 1910, 1930\}$ I build the cumulative rail network in place at the start of the wave, keeping all segments with an operational year at or before t . Seg-

ment vertices are snapped to a 1 km grid so that shared endpoints across adjacent segments are merged into common nodes, yielding a connected, routable graph in which edge weights equal true segment length. Each CZ centroid is linked by a straight-line “access edge” to its $K = 20$ nearest rail nodes, proxying for the unpaved wagon/road trip required to reach the railroad; no digitized historical road network is used for this leg. The grid resolution and K are computational choices made for tractability.³⁸

For every ordered CZ pair (o, d) and wave t , I compute the least-cost route on the combined graph, where travel on the rail network costs 1 per unit distance and travel off the network costs κ per unit distance:

$$\text{distance}_t(o, d) = \min_{\text{routes}} \left[\text{km}_{\text{on-rail}} \times 1 + \text{km}_{\text{off-rail}} \times \kappa \right], \quad (\text{A4.2.1})$$

solved by Dijkstra’s shortest-path algorithm from every CZ centroid. By construction, $\text{eff_dist}_t(o, d)$ is bounded between the Euclidean distance (perfect rail connection) and $\kappa \times$ Euclidean distance (no useful rail access, pure wagon travel), and is floored at the latter to handle pairs with no nearby rail. This yields a distinct effective-distance matrix for each wave, so that the same CZ pair becomes relatively closer over time as the rail network expands.

The parameter κ is the ratio of the per-mile cost of off-rail (wagon) transport to on-rail transport. Historical freight-rate evidence puts wagon haulage at roughly an order of magnitude more expensive per ton-mile than rail (Donaldson & Hornbeck, 2016; Fogel, 1964; Taylor, 1951); complementary hedonic evidence confirms rail access was capitalized into antebellum land values (Craig, Palmquist, & Weiss, 1998). Motivated by this range, I set $\kappa = 10$.³⁹

A4.3 Estimation of migration costs

I estimate migration costs from gross migration flows implied by the location-choice probabilities. Let F_{gnt}^j denote the mass of adults of ancestral type g and sector j who move from origin n to destination n' in period t . Given the location-choice probability in equation (13),

$$F_{gnt}^j = L_{gnt}^j P_t^N(n' | n, j, g), \quad (\text{A4.3.1})$$

where L_{gnt}^j is the mass of type- g , sector- j adults residing in origin n .

From equation (11), the location-choice probability can be written as

$$P_t^N(n' | n, j, g) = \exp \left\{ \frac{1}{\sigma_N} \left[\mathcal{A}_{gnt}^j - \rho \log L_{n't} + v(w_{n't}^j, p_{n't}) + \beta O_t(n', j, g) \right] \right\}$$

³⁸The size of K does not matter much for the effective distance with a median deviation of below 0.1%.

³⁹The results are robust to $\kappa \in \{3, 5, 10\}$.

$$+ \Theta_H H_{gnjt}(n') - \frac{M_{gnt}^j}{\sigma_N} - \log \sum_{n'' \in N} \exp\left(\frac{B_{gnt}^j(n'')}{\sigma_N}\right) \Bigg\}, \quad (\text{A4.3.2})$$

where

$$\Theta_H \equiv \frac{\theta_H}{\sigma_N}. \quad (\text{A4.3.3})$$

Substituting equation (A4.3.2) into equation (A4.3.1) gives

$$\begin{aligned} \mathbb{E}[F_{gnt}^j] = \exp \Bigg\{ & \log L_{gnt}^j - \log \sum_{n'' \in N} \exp\left(\frac{B_{gnt}^j(n'')}{\sigma_N}\right) \\ & + \frac{1}{\sigma_N} [\mathcal{A}_{gnt}^j - \rho \log L_{n't} + v(w_{n't}^j, p_{n't}) + \beta O_t(n', j, g)] + \Theta_H H_{gnjt}(n') - \frac{M_{gnt}^j}{\sigma_N} \Bigg\}. \end{aligned} \quad (\text{A4.3.4})$$

This expression separates naturally into origin- and destination-specific components. Define the destination component as

$$\text{destFE}_{gnt}^j \equiv \frac{1}{\sigma_N} [\mathcal{A}_{gnt}^j - \rho \log L_{n't} + v(w_{n't}^j, p_{n't}) + \beta O_t(n', j, g)], \quad (\text{A4.3.5})$$

which contains all components of destination utility that vary with g , n' , j and t , but not with the origin n .

Similarly, define the origin component as

$$\text{origFE}_{gnt}^j \equiv \log L_{gnt}^j - \log \sum_{n'' \in N} \exp \left[\text{destFE}_{gnt}^j + \Theta_H H_{gnjt}(n'') - \frac{M_{gnt}^j}{\sigma_N} \right]. \quad (\text{A4.3.6})$$

The first term is the mass of potential migrants at the origin, while the second is the location-choice inclusive value. Both are common across destinations for a given origin state (g, n, j, t) and can therefore be absorbed by an origin fixed effect.

Using equations (A4.3.5)–(A4.3.6), expected migration flows can be written as

$$\mathbb{E}[F_{gnt}^j] = \exp \left(\text{origFE}_{gnt}^j + \text{destFE}_{gnt}^j + \Theta_H H_{gnjt}(n') - \frac{M_{gnt}^j}{\sigma_N} \right). \quad (\text{A4.3.7})$$

Importantly, the kin term is not part of either fixed effect. The object $H_{gnjt}(n')$ depends jointly on the origin state and the destination: two adults of the same ancestral type and sector considering the same destination can have different kin exposure there depending on their origin. It therefore cannot in general be absorbed by separate origin and destination fixed effects. If not origin or destination fixed effect can be estimated for a group, I assign them the weighted average for a given destination or origin region by sector.

I parameterize normalized bilateral migration costs as

$$\frac{M_{gmn't}^j}{\sigma_N} = \lambda_t^j \tilde{d}_{nn't}, \quad (\text{A4.3.8})$$

where $\tilde{d}_{nn't}$ is the log of rail-adjusted least-cost distance between n and n' , normalized to zero for within-CZ moves. Substituting equation (A4.3.8) into equation (A4.3.7) yields the estimating equation

$$\mathbb{E}[F_{gmn't}^j] = \exp\left(\text{origFE}_{gnt}^j + \text{destFE}_{gn't}^j + \hat{\Theta}_H H_{gnjt}(n') - \lambda_t^j \tilde{d}_{nn't}\right). \quad (\text{A4.3.9})$$

I estimate equation (A4.3.9) by PPML (Silva & Tenreyro, 2006), using all origin-destination flows, including stayers. Because Θ_H is identified separately from the family-death design, I treat $\hat{\Theta}_H H_{gnjt}(n')$ as a known offset rather than re-estimating its coefficient. The coefficient λ_t^j therefore identifies distance-related migration costs net of kin exposure.

The estimated distance coefficient identifies migration costs relative to the location-choice scale. After estimating σ_N , I recover migration costs in utility units as

$$\hat{M}_{gmn't}^j = \hat{\sigma}_N \hat{\lambda}_t^j \tilde{d}_{nn't}. \quad (\text{A4.3.10})$$

Table A4.3.1 reports the estimated distance coefficients.

Table A4.3.1: Distance Elasticity of Migration Flows by Sector and Year

	1880		1910		1930	
	Agric. (1)	Manuf. (2)	Agric. (3)	Manuf. (4)	Agric. (5)	Manuf. (6)
Log(distance)	-1.152*** (0.013)	-1.144*** (0.018)	-1.222*** (0.015)	-1.135*** (0.022)	-1.406*** (0.019)	-1.290*** (0.022)
Observations	9,568,230	1,222,099	14,706,667	5,022,271	11,810,302	8,457,209
Pseudo R^2	0.559	0.542	0.591	0.529	0.675	0.626
Origin clusters	448	433	454	454	455	455
Destination clusters	448	444	454	454	455	455
Estimation	PPML	PPML	PPML	PPML	PPML	PPML

Notes: The dependent variable is the migration flow between origin and destination commuting zones for ancestry group g by origin sector. Each column reports a separate PPML regression by year and sector. All specifications include ancestry-group $\times$ origin and ancestry-group $\times$ destination fixed effects. Standard errors, reported in parentheses, are robust to heteroskedasticity and two-way clustered by origin and destination commuting zone. The constant is omitted from the table. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A4.4 Sector deaths and continuation-value adjustment

This appendix derives the mapping between the sector-death coefficients and the structural kin parameters and quantifies the error from ignoring the effect of a sector-employed relative on future location values. As in Section 5.2, $r = P$ denotes a parent death and $r = R$ an aunt-or-uncle death. For any object x ,

$$\Delta^{(r)}x \equiv x^{\text{post}(r)} - x^{\text{pre}}$$

denotes its post-death minus pre-death change following a death of type r .

Sector log odds. For sector k relative to the alternative sector $\ell \neq k$, the logit choice probabilities imply

$$\ell_{gnjt}^k \equiv \log \frac{P_t^O(k | n, j, g)}{P_t^O(\ell | n, j, g)} = \Lambda_k F_{gnjt}^k - \Lambda_\ell F_{gnjt}^\ell - \frac{\tau_{gnjt}^k - \tau_{gnjt}^\ell}{\sigma_O} + \frac{V_{t+1}(n, k, g) - V_{t+1}(n, \ell, g)}{\sigma_O}. \quad (\text{A4.4.1})$$

A death of a sector- k relative reduces the sector-access stock by

$$d_{F,P}^k = 1, \quad d_{F,R}^k = \psi_R^k.$$

The resulting change in sector- k log odds can therefore be written as

$$\Delta^{(r)}\ell^k = -\Lambda_k d_{F,r}^k + \Omega_r^k + \xi_r^k, \quad (\text{A4.4.2})$$

where

$$\Omega_r^k \equiv \frac{\Delta^{(r)}V_H(k) - \Delta^{(r)}V_H(\ell)}{\sigma_O} \quad (\text{A4.4.3})$$

captures the continuation-value response generated directly by the loss of geographic kin support. The residual term ξ_r^k collects changes in relative sector barriers and non-kin continuation values.

Continuation-value response. Let $d_{H,r}$ denote the reduction in location support caused by a death of type r . A parent contributes one unit, so $d_{H,P} = 1$; for an aunt or uncle, $d_{H,R}$ is the corresponding horizontal-location contribution implied by the kin block.

Consider a relative located at the origin n . Conditional on sector h , write the pre-death inclusive value as

$$V(h) = \sigma_N \log \left[\exp \left\{ \frac{B^h(n)}{\sigma_N} \right\} + \sum_{m \neq n} \exp \left\{ \frac{B^h(m)}{\sigma_N} \right\} \right]. \quad (\text{A4.4.4})$$

The death lowers origin utility by $\theta_H d_{H,r}$ while leaving the utility of alternative destinations

unchanged. The post-death inclusive value is therefore

$$V^{\text{post}(r)}(h) = \sigma_N \log \left[\exp \left\{ \frac{B^h(n)}{\sigma_N} \right\} e^{-\Theta_H d_{H,r}} + \sum_{m \neq n} \exp \left\{ \frac{B^h(m)}{\sigma_N} \right\} \right]. \quad (\text{A4.4.5})$$

Let

$$s_h \equiv P_{t+1}^N(n \mid n, h, g)$$

denote the pre-death probability of remaining at the origin. Factoring out the pre-death denominator gives

$$\Delta^{(r)} V_H(h) = \sigma_N \log [1 - s_h (1 - e^{-\Theta_H d_{H,r}})]. \quad (\text{A4.4.6})$$

Substituting this expression into equation (A4.4.3) yields

$$\Omega_r^k = \frac{\sigma_N}{\sigma_O} \log \left[\frac{1 - s_k (1 - e^{-\Theta_H d_{H,r}})}{1 - s_\ell (1 - e^{-\Theta_H d_{H,r}})} \right]. \quad (\text{A4.4.7})$$

The geographic component therefore cancels exactly when $s_k = s_\ell$. Otherwise,

$$\text{sign}(\Omega_r^k) = -\text{sign}(s_k - s_\ell). \quad (\text{A4.4.8})$$

For small Θ_H ,

$$\Omega_r^k \simeq -\frac{\sigma_N}{\sigma_O} \Theta_H d_{H,r} (s_k - s_\ell). \quad (\text{A4.4.9})$$

Thus the location channel affects the coefficient mapping only when the two sector choices imply different subsequent location behavior.

Identification and mapping error. Let $\tilde{D}_z^r$ denote death exposure residualized on the fixed effects and controls in equation (30). The identifying restriction is

$$\mathbb{E}[\tilde{D}_z^r \zeta_{r,z}^k] = 0. \quad (\text{A4.4.10})$$

This requires residual death exposure to be unrelated to shocks that independently change relative sector barriers or non-kin continuation values. Origin-CZ-by-parental-sector-by-year fixed effects absorb sectoral shocks common to workers in the same initial state, while ancestry fixed effects and predetermined controls absorb persistent differences across ancestral groups. Lagged sector-specific migration rates are also included and strongly predict subsequent migration rates, helping absorb persistent heterogeneity in future location behavior.

Under equation (A4.4.10), the regression coefficient satisfies

$$\beta_r^k = \Lambda_k d_{F,r}^k - \bar{\Omega}_r^k, \quad (\text{A4.4.11})$$

where $\bar{\Omega}_r^k$ is the continuation-value correction averaged over the identifying variation in the regression. Using the simpler mapping $\beta_r^k = \Lambda_k d_{F,r}^k$ therefore produces the error

$$\underbrace{\beta_r^k - \Lambda_k d_{F,r}^k}_{\text{mapping error}} = -\bar{\Omega}_r^k. \quad (\text{A4.4.12})$$

The corresponding corrected parameters are

$$\Lambda_k^{\text{corr}} = \beta_P^k + \bar{\Omega}_P^k, \quad \psi_R^{k,\text{corr}} = \frac{\beta_R^k + \bar{\Omega}_R^k}{\beta_P^k + \bar{\Omega}_P^k}. \quad (\text{A4.4.13})$$

Magnitude. The continuation-value adjustment is quantitatively small. Average probabilities of remaining at the origin are 0.734 for agricultural workers and 0.744 for manufacturing workers. Using the model-implied sector-specific location probabilities together with

$$\hat{\sigma}_N = 0.51, \quad \hat{\sigma}_O = 0.60, \quad \hat{\Theta}_H = 0.024, \quad \hat{\psi}_R^H = 0.874,$$

equation (A4.4.7) gives

$$\bar{\Omega}_P^M = -0.000205, \quad \bar{\Omega}_P^A = 0.000205,$$

and

$$\bar{\Omega}_R^M = -0.000179, \quad \bar{\Omega}_R^A = 0.000179.$$

Applying equation (A4.4.13) would change the parental sector parameters from

$$\Lambda_M = 0.0489 \quad \text{to} \quad \Lambda_M^{\text{corr}} = 0.0487,$$

and from

$$\Lambda_A = 0.0206 \quad \text{to} \quad \Lambda_A^{\text{corr}} = 0.0208.$$

These correspond to changes of approximately 0.4% and 1.0%, respectively. The implied horizontal relationship weights change from

$$\psi_R^M = 0.141 \quad \text{to} \quad \psi_R^{M,\text{corr}} = 0.138,$$

and from

$$\psi_R^A = 0.305 \quad \text{to} \quad \psi_R^{A,\text{corr}} = 0.311,$$

or by approximately 2.1% in either case. Equivalently, the absolute aunt-or-uncle continuation corrections are about 2.6% of the manufacturing aunt coefficient and 2.8% of the agricultural aunt coefficient.

Because the adjustment is small relative to the estimated sector-kin effects and leaves the

substantive conclusions unchanged, the baseline quantification uses the direct coefficient mapping in equation (31). Equation (A4.4.13) shows that explicitly accounting for the location continuation-value channel has negligible quantitative effects on the recovered kin parameters.

A4.5 Indirect Inference

This appendix provides the details of the indirect-inference procedure used to estimate the location- and sector-choice dispersion parameters (σ_N, σ_O) . I first describe the empirical Dust Bowl moments and the agricultural productivity disturbance, then explain how candidate dispersion parameters enter the model inversion, how I construct the corresponding Dust Bowl counterfactuals and how the empirical and simulated moments are matched.

A4.5.1 Empirical moments

The Dust Bowl generated large and persistent losses in agricultural land quality and economic activity across locations with different erosion severity (Hornbeck, 2012). These shocks induced substantial migration and longer-run adjustment (Hornbeck, 2012, 2023; Liu, 2026). I combine full-count Census microdata (Ruggles et al., 2024), longitudinal links from the IPUMS Multi-generational Longitudinal Panel (Helgertz et al., 2023) and the Census Tree (Buckles et al., 2025), soil-erosion measures from Hornbeck (2012) and local historical controls from Fishback (2017).

Let c index commuting zones and t the baseline census year, with outcomes measured twenty years later. Because erosion severity is time invariant, I interact medium- and high-erosion indicators with baseline-year indicators and estimate

$$Y_{i,c,t+20} = \alpha + \lambda_c + \lambda_{s(c),t} + \gamma' X_{i,c,t} + \sum_{r \neq 1910} \mathbf{1}\{t = r\} [\beta_{M,r} \text{MErosion}_c + \beta_{H,r} \text{HErosion}_c] + \varepsilon_{i,c,t+20}, \quad (\text{A4.5.1})$$

where 1910 is the baseline cohort, λ_c are CZ fixed effects and $\lambda_{s(c),t}$ are state-by-baseline-year fixed effects. Medium and high erosion are defined from cumulative topsoil loss, with low-erosion CZs omitted (Hornbeck, 2012). The controls $X_{i,c,t}$ contain predetermined individual and household characteristics and local Depression-era controls. Standard errors are clustered by CZ.

For sector choice, the sample consists of children younger than 15 at baseline and the outcome is manufacturing employment twenty years later, estimated separately by parental sector.⁴⁰ For migration, the sample consists of adults older than 25 and the outcome indicates whether the

⁴⁰Manufacturing contains durable industries (IND1950 306–399) and non-durable industries (IND1950 406–499). Agriculture is farming (IND1950 105).

individual moves across commuting zones, estimated separately by initial sector. The resulting erosion-by-cohort coefficients correspond to changes in the sector-choice and out-migration probabilities generated by the model.

Let $\widehat{b}^{\text{data}}$ collect the empirical coefficients used as indirect-inference moments. The same samples, erosion definitions and auxiliary specifications are applied to simulated model outcomes below.

A4.5.2 Productivity estimates

The indirect-inference exercise also requires the magnitude of the agricultural productivity disturbance that generates these behavioral responses. Rather than mapping agricultural revenue per acre into productivity through an auxiliary production function, I estimate the erosion effect directly on the agricultural productivity shifter Z_{ct}^A recovered from the quantitative model.

Agricultural productivity is recovered for each CZ and model period from the goods-market inversion described in Appendix A4.6.1. This inversion combines observed wages and employment with trade costs, expenditure shares and the Eaton–Kortum trade structure to recover Z_{ct}^A . Importantly, this recovery does not use the location- and sector-choice dispersion parameters (σ_N, σ_O) subsequently estimated by indirect inference.

I estimate

$$\log Z_{ct}^A = \alpha_c + \alpha_{s(c),t} + \sum_{\tau \in \{1910, 1950\}} [\beta_{M,\tau} \text{MErosion}_c + \beta_{H,\tau} \text{HErosion}_c] \mathbf{1}\{t = \tau\} + \gamma' X_{ct} + \varepsilon_{ct}, \quad (\text{A4.5.2})$$

where 1930 is the omitted year. CZ fixed effects absorb permanent differences across locations and state-by-year fixed effects absorb shocks common to CZs within a state. The preferred specification additionally includes the local controls reported in Table A4.5.2. Standard errors are clustered by CZ.

Columns (3)–(4) of Table A4.5.2 report the estimates. The pre-Dust-Bowl coefficients are small and statistically indistinguishable from zero: in the preferred specification, the 1910 coefficients are 0.008 for medium erosion and -0.019 for high erosion. By 1950, agricultural productivity is 0.179 log points lower in medium-erosion CZs and 0.226 log points lower in high-erosion CZs relative to low-erosion CZs.

I use these estimates to construct the Dust Bowl component of agricultural productivity,

$$\widehat{\Delta \log Z_{c,1950}^{DB}} = \widehat{\beta}_{M,1950} \text{MErosion}_c + \widehat{\beta}_{H,1950} \text{HErosion}_c. \quad (\text{A4.5.3})$$

The observed transition retains this component. In the no-Dust-Bowl counterfactual used for

Table A4.5.1: Erosion exposure, sector choice and migration

	Children			Household heads	
	Farmer children		Non-farmer children	Migration	
	Agriculture (1)	Manufacturing (2)	Agriculture (3)	Farmer (4)	Non-farmer (5)
<i>Panel A. Children's sector choice</i>					
<i>Pre-trends</i>					
1900 × Medium erosion	1.017 (0.628)	-0.910* (0.532)	-0.314 (0.404)		
1900 × High erosion	0.617 (0.772)	-0.664 (0.569)	-0.342 (0.500)		
<i>Post-treatment</i>					
1930 × Medium erosion	-3.87*** (1.05)	2.27*** (0.82)	-0.71 (0.51)		
1930 × High erosion	-6.23*** (1.30)	3.26*** (1.15)	-0.97 (0.64)		
<i>Panel B. Household-head migration</i>					
<i>Pre-trends</i>					
1900 × Medium erosion				-0.755 (0.887)	-0.639 (0.425)
1910 × Medium erosion				-0.354 (0.669)	0.241 (0.343)
1900 × High erosion				-0.671 (1.011)	0.633 (0.446)
1910 × High erosion				-0.945 (0.846)	-0.222 (0.342)
<i>Post-treatment</i>					
1930 × Medium erosion				1.12*** (0.40)	1.00*** (0.37)
1930 × High erosion				0.93** (0.42)	0.17 (0.46)
Observations	4,112,128	4,112,128	7,541,146	4,441,056	19,507,083
R^2	0.109	0.055	0.063	0.045	0.043
Mean outcome in 1910	51.00	23.00	8.00	5.00	8.00
Mean medium-erosion exposure	44.00	44.00	41.00	44.00	40.00
Mean high-erosion exposure	19.00	19.00	11.00	18.00	11.00
CZ fixed effects	Yes	Yes	Yes	Yes	Yes
State × year fixed effects	Yes	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes	Yes

Notes: The table reports difference-in-differences estimates relating medium and high erosion exposure to children's sector choices and household-head migration. Columns (1)–(2) restrict the sample to children of farmers; column (3) uses children of non-farmers. Columns (4)–(5) report migration outcomes for farmer and non-farmer household heads, respectively. Coefficients are multiplied by 100 and are therefore expressed in percentage points. All specifications include commuting-zone fixed effects, state-by-year fixed effects and individual controls. Standard errors, clustered by commuting zone, are reported in parentheses. The omitted erosion category is low exposure. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

indirect inference, I remove it from agricultural productivity:

$$\log Z_{c,1950}^{A,\text{noDB}} = \log Z_{c,1950}^{A,\text{obs}} - \widehat{\Delta \log Z_{c,1950}^{DB}}. \quad (\text{A4.5.4})$$

Because the estimated Dust Bowl components are negative, removing them raises agricultural productivity in medium- and high-erosion locations relative to the observed economy.

Comparison with agricultural outcomes in the data. As an additional validation of the model-recovered productivity disturbance, I estimate the same event-study specification using log farm products per acre. Measures of agricultural production or revenue per acre are standard outcomes in empirical work on the Dust Bowl; related measures also enter the quantitative analysis in [Yang \(2024\)](#).

Columns (1)–(2) of Table [A4.5.2](#) show that farm products per acre decline by 0.176 log points in medium-erosion CZs and 0.241 log points in high-erosion CZs by 1950. These magnitudes are close to the corresponding model-recovered productivity effects of 0.179 and 0.226 log points. The agreement is not imposed by the productivity inversion and provides an external check that the erosion-related decline in Z^A is of a magnitude consistent with the deterioration observed directly in agricultural production.

A4.5.3 Candidate-specific model inversion

The empirical responses identify how strongly workers react to the Dust Bowl, while the productivity estimates identify the size of the disturbance generating that response. For each candidate pair

$$\sigma = (\sigma_N, \sigma_O),$$

I recover the remaining structural costs and amenities using the model inversion in [Appendix A4.6](#).

The kin parameters identified earlier are expressed relative to the corresponding taste dispersion. For a candidate σ_N , the estimated location-attachment coefficient and migration-cost schedule are mapped into utility units according to

$$\theta_H = \sigma_N \Theta_H, \quad M_{gnmt}^j = \sigma_N \lambda_t^j \tilde{d}_{nmt}.$$

Given these objects and observed location-choice probabilities, the location inversion nets out kin amenities and bilateral migration costs to recover the non-kin location amenities that rationalize observed migration choices.

Similarly, for sector choice, $\theta_k = \sigma_O \Lambda_k$. Given the candidate σ_O , observed sector-choice odds and the corresponding continuation values, I recover the non-kin sector-entry costs after netting out the kin contribution Q_{gnjt}^k . Because continuation values themselves depend on σ_N and the

Table A4.5.2: Estimating the effect of soil erosion on agricultural productivity

	Data		Model	
	Log farm products per acre (1)	(2)	Log agricultural productivity (3)	(4)
<i>Panel A. Medium versus low erosion</i>				
1910 × Medium erosion	0.130 (0.116)	0.130 (0.117)	0.001 (0.059)	0.008 (0.060)
1950 × Medium erosion	-0.176* (0.096)	-0.176* (0.098)	-0.186 (0.117)	-0.179 (0.117)
1970 × Medium erosion	-0.053 (0.128)	-0.053 (0.129)		
1990 × Medium erosion	0.004 (0.148)	0.004 (0.150)		
<i>Panel B. High versus low erosion</i>				
1910 × High erosion	0.273** (0.128)	0.273** (0.128)	-0.036 (0.054)	-0.019 (0.056)
1950 × High erosion	-0.241*** (0.092)	-0.241** (0.094)	-0.231* (0.118)	-0.226* (0.118)
1970 × High erosion	-0.268* (0.140)	-0.268* (0.142)		
1990 × High erosion	-0.446** (0.187)	-0.446** (0.190)		
Observations	1,857	1,857	1,665	1,647
R^2	0.849	0.849	0.939	0.939
CZ fixed effects	✓	✓	✓	✓
State × year fixed effects	✓	✓	✓	✓
Log population		✓		✓
Years	1910–1990	1910–1990	1910–1950	1910–1950
Cluster	CZ	CZ	CZ	CZ

Notes: The unit of observation is a CZ by year. Columns (1)–(2) use the log value of farm products per acre as the measure of observed agricultural productivity. Columns (3)–(4) use agricultural productivity recovered from the quantitative model. Medium- and high-erosion CZs are compared with low-erosion CZs, with 1930 as the omitted year. All specifications include CZ and state-by-year fixed effects. Columns (2) and (4) additionally control for log population. Standard errors, clustered at the CZ level, are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

equilibrium path, the recovered sector-entry costs are candidate specific.

Thus, each candidate (σ_N, σ_O) implies a complete set of fundamentals that exactly rationalizes the observed baseline choices while preserving the kin effects identified relative to taste dispersion from the empirical kin variation. This inversion is repeated before evaluating the Dust Bowl response for that candidate.

A4.5.4 Model-based Dust Bowl effects

For each candidate σ , I solve two equilibrium transitions using the candidate-specific recovered fundamentals.

Observed transition (Dust Bowl). The first transition follows the observed economy and includes the estimated Dust Bowl component of agricultural productivity. Sector choices, migration, wages, prices and the kin distribution respond endogenously. Because observed fundamentals end in 1950, I hold them fixed thereafter for the final transition used to evaluate continuation values.

Counterfactual transition (no Dust Bowl). The second transition holds all non-Dust-Bowl fundamentals and recovered wedges fixed but removes the estimated erosion-induced component of agricultural productivity in equation (A4.5.4). I then resolve the model for the counterfactual equilibrium path.

Let $Y_{i,c,t+20}^{\text{DB}}(\sigma)$ and $Y_{i,c,t+20}^{\text{noDB}}(\sigma)$ denote the corresponding model-implied outcomes. Their difference,

$$\tau_{i,c,t+20}^{\text{model}}(\sigma) = Y_{i,c,t+20}^{\text{DB}}(\sigma) - Y_{i,c,t+20}^{\text{noDB}}(\sigma), \quad (\text{A4.5.5})$$

isolates the model-implied Dust Bowl response while allowing sector choice, migration, wages, prices and kin connections to adjust in equilibrium. All non-Dust-Bowl fundamentals are held fixed across the two paths.

A4.5.5 Simulated regressions and indirect inference

To make the simulated moments directly comparable with the empirical estimands, I apply the same auxiliary design used in equation (A4.5.1) to the model-generated outcomes. For each candidate $\sigma = (\sigma_N, \sigma_O)$, I aggregate model outcomes to the corresponding ancestry-group-by-CZ cells and estimate

$$\begin{aligned} Y_{g,c,t+20}^{\text{model}}(\sigma) = & \alpha^{\text{model}}(\sigma) + \lambda_c^{\text{model}}(\sigma) + \lambda_{s(c),t}^{\text{model}}(\sigma) \\ & + \sum_{r \neq 1910} \mathbf{1}\{t = r\} [\beta_{M,r}^{\text{model}}(\sigma) \text{MErosion}_c + \beta_{H,r}^{\text{model}}(\sigma) \text{HErosion}_c] + u_{g,c,t+20}^{\text{model}}(\sigma). \end{aligned} \quad (\text{A4.5.6})$$

I use the same omitted cohort and fixed-effect structure as in the empirical auxiliary regression.⁴¹ Let $b^{\text{model}}(\sigma)$ collect the same erosion-by-cohort coefficients from equation (A4.5.6) that enter

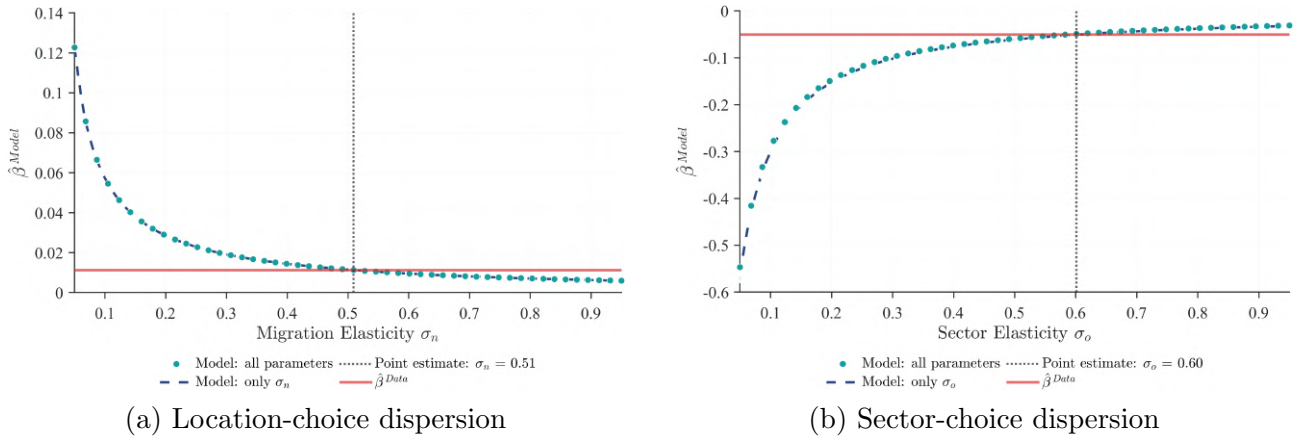
⁴¹The empirical regression is estimated at the individual level, whereas the simulated regression uses ancestry-group-by-CZ cells generated by the model.

$\hat{b}^{\text{data}}$. The indirect-inference estimator is

$$\hat{\sigma} = \arg \min_{\sigma \in \Sigma} m(\sigma)' W m(\sigma), \quad m(\sigma) = b^{\text{model}}(\sigma) - \hat{b}^{\text{data}}, \quad W = I. \quad (\text{A4.5.7})$$

The identity weighting matrix gives equal weight to the targeted auxiliary coefficients. The current implementation evaluates the objective on a 50×50 grid over (σ_N, σ_O) . Figure A4.5.1 reports the resulting fit.

Figure A4.5.1: Indirect inference results



Notes: Panel (a) shows how the model-implied migration response to the Dust Bowl varies with the location-choice dispersion parameter σ_N . Panel (b) shows the corresponding sectoral response as a function of the sector-choice dispersion parameter σ_O . The selected pair minimizes equation (A4.5.7).

A4.5.6 Timing and identifying assumptions

The indirect-inference exercise relies on four maintained assumptions.

First, I treat the Dust Bowl as unanticipated in 1930. Once the shock occurs, agents understand the subsequent productivity path used in the transition. This prevents pre-1930 choices from responding to a shock that had not yet occurred.

Second, observed fundamentals are unavailable after 1950. I therefore hold the 1950 fundamentals fixed for the final model period used to evaluate continuation values. This terminal assumption affects future values but does not introduce an additional post-1950 shock.

Third, erosion affects workers through agricultural productivity rather than directly changing migration costs or non-kin sector-entry costs. Baseline farm-ownership controls, heterogeneity by ownership status and alternative specifications provide checks on this restriction (Liu, 2026; Yang, 2024).

Fourth, the no-Dust-Bowl counterfactual removes the estimated productivity disturbance symmetrically while holding other recovered fundamentals fixed. The exercise therefore abstracts

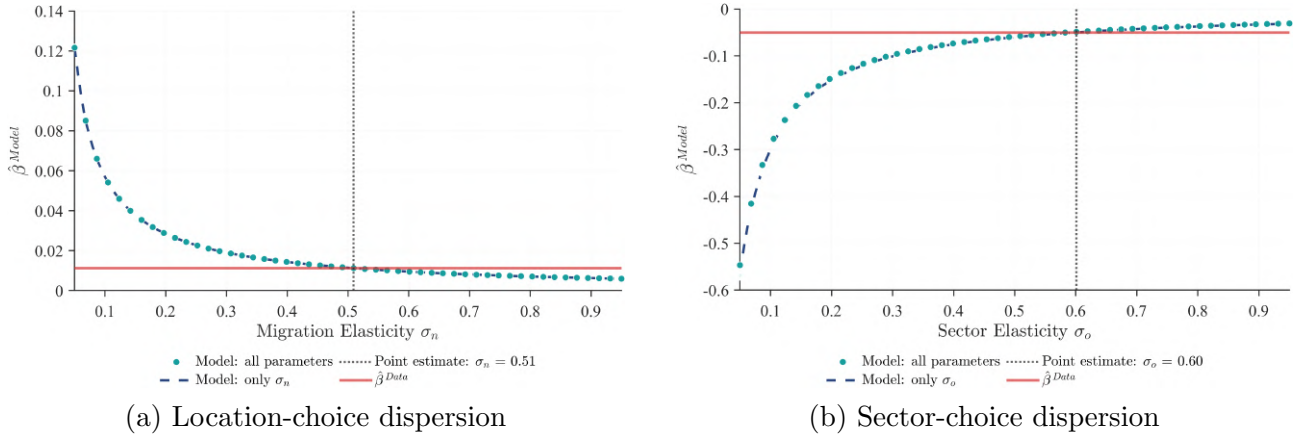
from additional state-dependent responses, such as borrowing constraints, that would themselves change because of the shock.

Under these assumptions, the variation identifying (σ_N, σ_O) comes from the response of migration and sector choice to an independently measured agricultural productivity disturbance rather than from cross-sectional differences in the levels of those choices.

A4.5.7 Indirect inference without social ties

For the counterfactual analysis without social ties, I re-estimate the dispersion parameters because their minimizing values could in principle change when the kin mechanism is removed. Figure A4.5.2 documents the results. In the current implementation, the minimizing values are unchanged for both σ_N and σ_O .

Figure A4.5.2: Indirect inference results without social ties



Notes: Panel (a) shows the model-implied migration response as a function of σ_N in the model without social ties. Panel (b) shows the corresponding sectoral response as a function of σ_O .

A4.6 Model inversion

A4.6.1 Productivity

This subsection describes how sector–location productivity is recovered from observed wages, employment, trade costs and expenditure shares. The inversion follows the logic of Takeda (2022) and combines labor-market clearing with the Eaton–Kortum trade structure and the PIGL non-homothetic demand system used in the main text (Boppart, 2014; Eckert & Peters, 2022).

For each sector $k \in \{A, M\}$ and location n , labor-market clearing requires that total labor

income equals total expenditure on varieties produced in that location,

$$w_{nt}^k L_{nt}^k = \sum_{i \in N} \pi_{int}^k X_{it}^k, \quad (\text{A4.6.1})$$

where w_{nt}^k is the sector- k wage in location n , L_{nt}^k is employment, π_{int}^k is the share of expenditure in destination i on varieties sourced from origin n and X_{it}^k is total expenditure in location i on sector k .

Expenditure is determined by the PIGL demand system. An individual in location i and sector j with nominal labor income w_{it}^j spends a share $\delta_k(w_{it}^j, p_{it})$ on good k , where from (5) the agricultural share is

$$\delta_A(y, p_{it}) = \varphi + \nu \left(\frac{y}{(p_{it}^A)^\varphi (p_{it}^M)^{1-\varphi}} \right)^{-\epsilon}. \quad \delta_M = 1 - \delta_A, \quad (\text{A4.6.2})$$

Because $\epsilon > 0$, the agricultural share falls with real income, so richer locations spend relatively less on agriculture—the non-homothetic Engel force that drives structural change. Aggregating across the two sectors of workers in location i , total sector- k expenditure is

$$X_{it}^k = \delta_k(w_{it}^A, p_{it}) w_{it}^A L_{it}^A + \delta_k(w_{it}^M, p_{it}) w_{it}^M L_{it}^M. \quad (\text{A4.6.3})$$

where each sector's workers spend out of their own labor income at their own PIGL shares.

Trade shares follow the Eaton–Kortum structure,

$$\pi_{int}^k = \frac{(w_{nt}^k T_{int}^k)^{-\mu_k} (Z_{nt}^k)^{\mu_k}}{\sum_{r \in N} (w_{rt}^k T_{irt}^k)^{-\mu_k} (Z_{rt}^k)^{\mu_k}} \quad (\text{A4.6.4})$$

with iceberg trade costs T_{int}^k , trade elasticity μ_k and unit cost x_{it}^k ; under the linear production technology unit costs equal wages, $x_{it}^k = w_{it}^k$. Substituting (A4.6.3) and (A4.6.4) into (A4.6.1) gives

$$w_{nt}^k L_{nt}^k = \sum_{i \in N} \frac{(x_{nt}^k T_{int}^k)^{-\mu_k} (Z_{nt}^k)^{\mu_k}}{\sum_{j \in N} (x_{jt}^k T_{ijt}^k)^{-\mu_k} (Z_{jt}^k)^{\mu_k}} [\delta_k(w_{it}^A, p_{it}) w_{it}^A L_{it}^A + \delta_k(w_{it}^M, p_{it}) w_{it}^M L_{it}^M]. \quad (\text{A4.6.5})$$

Factoring out the origin-specific productivity and unit-cost terms,

$$\begin{aligned} w_{nt}^k L_{nt}^k &= (x_{nt}^k)^{-\mu_k} (Z_{nt}^k)^{\mu_k} \sum_{i \in N} \frac{(T_{int}^k)^{-\mu_k}}{\sum_{j \in N} (x_{jt}^k T_{ijt}^k)^{-\mu_k} (Z_{jt}^k)^{\mu_k}} \\ &\quad \times [\delta_k(w_{it}^A, p_{it}) w_{it}^A L_{it}^A + \delta_k(w_{it}^M, p_{it}) w_{it}^M L_{it}^M], \end{aligned} \quad (\text{A4.6.6})$$

so productivity solves

$$(Z_{nt}^k)^{\mu_k} = w_{nt}^k L_{nt}^k (x_{nt}^k)^{\mu_k} \left(\sum_{i \in N} \frac{(T_{int}^k)^{-\mu_k}}{\sum_{j \in N} (x_{jt}^k T_{ijt}^k)^{-\mu_k} (Z_{jt}^k)^{\mu_k}} \times [\delta_k(w_{it}^A, p_{it}) w_{it}^A L_{it}^A + \delta_k(w_{it}^M, p_{it}) w_{it}^M L_{it}^M] \right)^{-1}. \quad (\text{A4.6.7})$$

To express the inversion in observable prices, use the Eaton–Kortum sectoral price equation $p_{nt}^k = \Gamma^k \left(\sum_{j \in N} (x_{jt}^k T_{njt}^k)^{-\mu_k} (Z_{jt}^k)^{\mu_k} \right)^{-1/\mu_k}$, which implies

$$\left(\frac{p_{it}^k}{\Gamma^k} \right)^{-\mu_k} = \sum_{j \in N} (x_{jt}^k T_{ijt}^k)^{-\mu_k} (Z_{jt}^k)^{\mu_k}. \quad (\text{A4.6.8})$$

Substituting (A4.6.8) into (A4.6.7) yields the closed-form productivity expression

$$Z_{nt}^k = (w_{nt}^k L_{nt}^k)^{1/\mu_k} x_{nt}^k \left(\sum_{i \in N} (T_{int}^k)^{-\mu_k} \left(\frac{p_{it}^k}{\Gamma^k} \right)^{\mu_k} \times [\delta_k(w_{it}^A, p_{it}) w_{it}^A L_{it}^A + \delta_k(w_{it}^M, p_{it}) w_{it}^M L_{it}^M] \right)^{-1/\mu_k}. \quad (\text{A4.6.9})$$

The system (A4.6.9) is solved by fixed-point iteration: starting from a guess for $\{Z_{nt}^k\}$, compute prices from the Eaton–Kortum price equation, form the PIGL expenditure shares (A4.6.2), evaluate the right-hand side and update productivity until convergence.

A4.6.2 Sector-entry costs and location amenities

This subsection describes how I invert observed sector and location choices into the remaining residual objects of the model: time-varying sector-entry costs and location-sector amenities. The kin parameters are estimated before the inversion. Hence, conditional on a candidate pair of choice scales (σ_N, σ_O) , the value of sector-specific kin connections and geographically proximate kin support are known. The inversion therefore treats the kin terms as structural components of the same choice problems used in the quantitative model rather than as additional reduced-form controls.

Step 0: inputs, kin objects and normalizations. For each date t , ancestral-origin type g , current location n , candidate destination m , current sector j and child sector k , the inversion takes as inputs the observed sector-choice probabilities $P_t^O(k \mid n, j, g)$, the migration flows used in the gravity estimation, employment, wages, prices, the current kin state, normalized migration costs and the scale-free kin parameters. The choice scales are not fixed prior to

indirect inference. Instead, for each candidate $\sigma = (\sigma_N, \sigma_O)$, I perform the inversion conditional on that candidate. The final reported fundamentals are evaluated at $\hat{\sigma} = (\hat{\sigma}_N, \hat{\sigma}_O)$.

For an adult of ancestral type g currently in location n and sector j , define

$$\Phi_{gnjt}^k(m) \equiv \Phi_{g,t}(n, j; m, k), \quad (\text{A4.6.10})$$

so that $\Phi_{gnjt}^k(m)$ is the expected mass of horizontal kin located in m and employed in sector k . The sector-specific kin stock entering the child's sector choice is

$$F_{gnjt}^k = \mathbf{1}\{j = k\} + \psi_R^k \Phi_{gnjt}^k(n), \quad (\text{A4.6.11})$$

because only same-region, same-sector horizontal kin enter the baseline sector-access stock. Conditional on the candidate sector-choice scale,

$$Q_{gnjt}^k = \theta_k F_{gnjt}^k, \quad \theta_k = \sigma_O \Lambda_k. \quad (\text{A4.6.12})$$

The geographically provided kin stock entering location choice is

$$H_{gnjt}(m) = \mathbf{1}\{m = n\} + \psi_R^H \sum_{\ell \in \{A, M\}} \Phi_{gnjt}^\ell(m), \quad (\text{A4.6.13})$$

and its utility value is

$$\theta_H H_{gnjt}(m), \quad \theta_H = \sigma_N \Theta_H. \quad (\text{A4.6.14})$$

Thus, the parent is normalized to contribute one unit in both kin margins. Horizontal kin receive the estimated relationship-specific weights ψ_R^k and ψ_R^H . In the baseline model,

$$Q_{gnjt}^k = \theta_k F_{gnjt}^k.$$

The same inversion applies to the schooling extension in Appendix A3.2.3 after replacing Q_{gnjt}^k by the corresponding schooling-augmented sector-entry value.

Sector-choice probabilities identify only relative sector utilities. For each parental sector j , I normalize the cost of remaining in the parent's sector to zero:

$$\tau_{gnjt}^j = 0 \quad \text{for all } (g, n, j, t). \quad (\text{A4.6.15})$$

The time-varying cost of entering the alternative sector $k \neq j$ is then identified from relative sector-choice probabilities. Thus, at each date the model recovers the two directional switching costs τ_{gnAt}^M and τ_{gnMt}^A .

Location-choice probabilities likewise identify destination values only up to an additive constant.

I use the same destination-value normalization in every step of the inversion and preserve that normalization across counterfactuals. The normalization affects the levels assigned to residual amenities but not the location probabilities generated by the model.

Step 1: recover normalized destination values and adult continuation values. The migration-gravity equation in Appendix A4.3 is estimated with the kin term $\hat{\Theta}_H H_{gnjt}(m)$ included as a known offset. It is therefore useful to define the non-kin destination value

$$\mathcal{D}_{gmt}^j \equiv \mathcal{A}_{gmt}^j - \rho \log L_{mt} + v(w_{mt}^j, p_{mt}) + \beta O_t(m, j, g). \quad (\text{A4.6.16})$$

The structural migration-flow equation can consequently be written as

$$\mathbb{E}[F_{gnmt}^j] = \exp \left[\text{orig}_{gmt}^j + \frac{\mathcal{D}_{gmt}^j}{\sigma_N} + \Theta_H H_{gnjt}(m) - \lambda_t^j \tilde{d}_{nmt} \right]. \quad (\text{A4.6.17})$$

The destination fixed effect recovered from the PPML regression therefore satisfies

$$\text{dest}_{gmt}^j = \frac{\mathcal{D}_{gmt}^j}{\sigma_N}, \quad (\text{A4.6.18})$$

up to the destination-value normalization imposed in the gravity estimation. Conditional on a candidate σ_N , the destination value in utility units is therefore

$$\mathcal{D}_{gmt}^j = \sigma_N \text{dest}_{gmt}^j. \quad (\text{A4.6.19})$$

The same result can be seen directly from migration odds. For any two candidate destinations m and r ,

$$\begin{aligned} \mathcal{D}_{gmt}^j - \mathcal{D}_{grt}^j &= \sigma_N \log \frac{P_t^N(m \mid n, j, g)}{P_t^N(r \mid n, j, g)} \\ &\quad - \theta_H [H_{gnjt}(m) - H_{gnjt}(r)] + M_{gnmt}^j - M_{grnt}^j. \end{aligned} \quad (\text{A4.6.20})$$

Thus, once bilateral migration costs and kin support are removed, destination odds recover the relative non-kin destination value.

Given the normalized destination values, the adult continuation value follows directly from the location-choice log sum:

$$V_t(n, j, g) = \sigma_N \log \sum_{m \in N} \exp \left[\frac{\mathcal{D}_{gmt}^j + \theta_H H_{gnjt}(m) - M_{gnmt}^j}{\sigma_N} \right]. \quad (\text{A4.6.21})$$

Using $\theta_H = \sigma_N \Theta_H$, $M_{gnmt}^j = \sigma_N \lambda_t^j \tilde{d}_{nmt}$, the same expression can be written entirely in normal-

ized units as

$$\frac{V_t(n, j, g)}{\sigma_N} = \log \sum_{m \in N} \exp \left[\text{dest}_{gmt}^j + \Theta_H H_{gnjt}(m) - \lambda_t^j \tilde{d}_{nmt} \right]. \quad (\text{A4.6.22})$$

The origin fixed effect provides an equivalent representation. Because the gravity equation is estimated on flows,

$$\text{orig}_{gnt}^j = \log L_{gnt}^j - \frac{V_t(n, j, g)}{\sigma_N}, \quad (\text{A4.6.23})$$

under the same fixed-effect normalization. Hence,

$$V_t(n, j, g) = \sigma_N [\log L_{gnt}^j - \text{orig}_{gnt}^j]. \quad (\text{A4.6.24})$$

Equations (A4.6.22) and (A4.6.24) therefore provide two equivalent ways of recovering the adult continuation value from the migration block. I use (A4.6.21).

Step 2: recover time-varying sector-entry costs. A child whose parent is in state (n, j) chooses sector k before making the next-period adult location decision. The systematic sector-choice value is

$$U_{gnjt}^k = Q_{gnjt}^k - \tau_{gnjt}^k + V_{t+1}(n, k, g). \quad (\text{A4.6.25})$$

Importantly, once the child chooses k , that sector becomes the child's own sector state. The continuation value is therefore $V_{t+1}(n, k, g)$, not a continuation value indexed by the parent's sector j .

With $\tau_{gnjt}^j = 0$, the sector-choice log odds imply, for the alternative sector $k \neq j$,

$$\tau_{gnjt}^k = [Q_{gnjt}^k + V_{t+1}(n, k, g)] - [Q_{gnjt}^j + V_{t+1}(n, j, g)] - \sigma_O \log \frac{P_t^O(k | n, j, g)}{P_t^O(j | n, j, g)}. \quad (\text{A4.6.26})$$

Observed sector-choice odds therefore identify the residual cost of entering the alternative sector at date t , after accounting for inherited sector-specific kin access and differences in future location values across sectors.

The inversion is performed separately at each date. For parents initially employed in agriculture, equation (A4.6.26) recovers τ_{gnAt}^M , the cost of entering manufacturing. For parents initially employed in manufacturing, it recovers τ_{gnMt}^A , the cost of entering agriculture. These residual costs are allowed to evolve over time and therefore absorb changes in sector-entry barriers not otherwise captured by the observed kin state, continuation values or other explicit model components.

Step 3: compute children's opportunity values. Given the recovered time-varying sector-entry costs, sector-specific kin values and adult continuation values, the opportunity

value available to the next generation is

$$O_t(n, j, g) = \sigma_O \log \sum_{k \in \{A, M\}} \exp \left[\frac{Q_{gnjt}^k - \tau_{gnjt}^k + V_{t+1}(n, k, g)}{\sigma_O} \right]. \quad (\text{A4.6.27})$$

This is the same sector-choice inclusive value that enters the parent's location problem through altruism. The object therefore links the sector and location blocks without introducing an additional reduced-form continuation term.

Step 4: recover location-sector amenities. Once $O_t(m, j, g)$ has been recovered, I can decompose the destination value in equation (A4.6.16). Rearranging gives

$$\mathcal{A}_{gmt}^j = \mathcal{D}_{gmt}^j + \rho \log L_{mt} - v(w_{mt}^j, p_{mt}) - \beta O_t(m, j, g). \quad (\text{A4.6.28})$$

The recovered amenity inherits the additive normalization of the destination fixed effects. This normalization is held fixed throughout the quantitative exercises. Counterfactual outcomes depend on utility differences and are therefore invariant to a common additive renormalization applied consistently to the destination values and the associated residual objects.

Step 5: implementation and terminal values. The inversion is conditional on a candidate pair

$$(\sigma_N, \sigma_O)$$

inside the indirect-inference procedure. For each candidate, I first convert the scale-free objects estimated in the kin and gravity blocks into utility units:

$$\theta_H = \sigma_N \Theta_H, \quad \theta_k = \sigma_O \Lambda_k, \quad M_{gmt}^j = \sigma_N \lambda_t^j \tilde{d}_{nmt}, \quad \mathcal{D}_{gmt}^j = \sigma_N \text{dest}_{gmt}^j. \quad (\text{A4.6.29})$$

I then compute $V_t(n, j, g)$ at each observed date directly from equation (A4.6.22). Thus, the observed location-choice block supplies the adult continuation values along the historical path. Sector choice at the final observed date T requires V_{T+1} , which is not directly recovered from an observed $T + 1$ migration distribution. I obtain this terminal continuation value from the terminal extension used in the quantitative model during the counterfactuals. For the amenity inversion, I assume that the amenities and sectoral barriers are the same in 1970 as they are in 1950.

Conditional on these continuation values, for each observed date t I perform the following steps:

1. recover the two directional time-varying sector-entry costs, τ_{gnAt}^M and τ_{gnMt}^A , from equation (A4.6.26);
2. compute $O_t(n, j, g)$ from equation (A4.6.27);

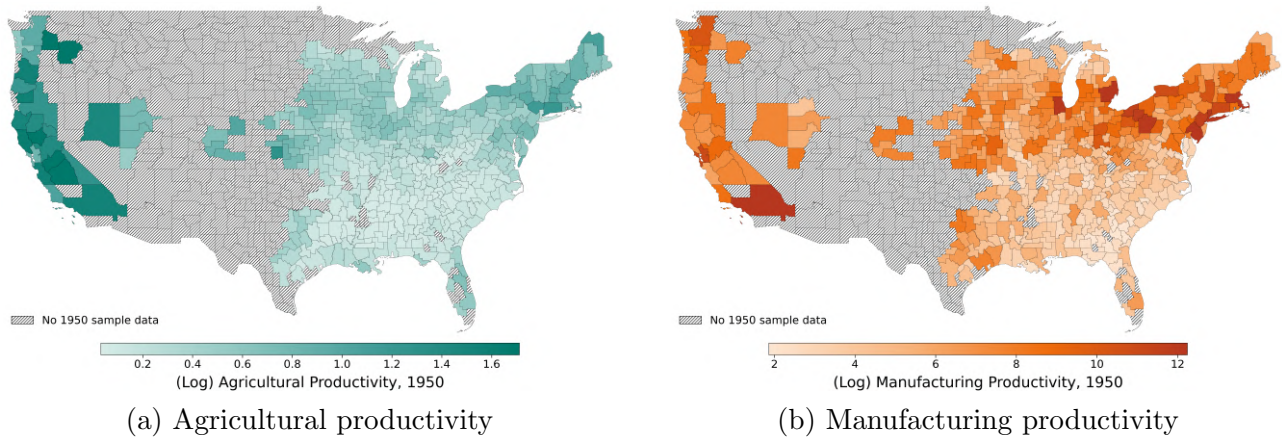
3. recover $\mathcal{A}_{gmt}^j$ from equation (A4.6.28).

A4.6.3 Model inversion: additional results

This appendix reports additional inversion results that complement the main validation exercise. I report the geography of recovered fundamentals, additional untargeted validation moments, the evolution of sectoral costs and the transition in family input shares.

Geography of recovered fundamentals. The inversion recovers, for each commuting zone and year, sectoral productivities, sectoral amenities and market-support prices. These objects are not targeted on external spatial series. Figure A4.6.1 maps recovered productivity in 1950. Agricultural productivity is highest in the West and Great Plains and lowest in the industrial Northeast, while manufacturing productivity is concentrated in the Northeast and around the Great Lakes. Figure A4.6.2 maps recovered amenities in 1880. These objects vary systematically across space without being directly targeted in the inversion.

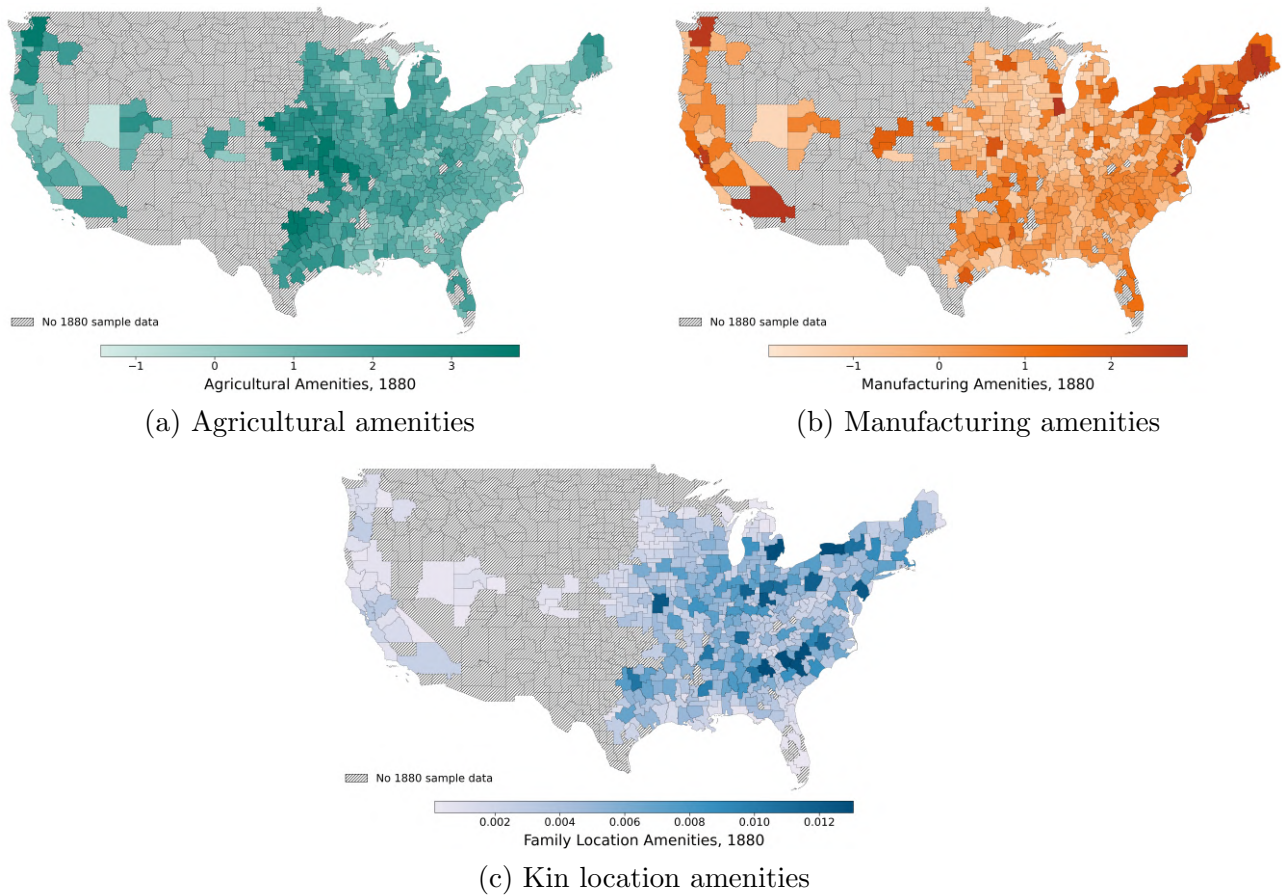
Figure A4.6.1: Geographical distribution of productivity, 1950



Notes: Recovered log sectoral productivity across commuting zones in 1950.

Additional untargeted validation. Figure A4.6.3 reports the full productivity validation. Recovered agricultural productivity aligns with agricultural output per worker in both 1880 and 1950, although the output-per-worker series is not used in the inversion. Figure A4.6.4 reports the corresponding employment-growth validation for both sectors. The model reproduces the spatial dynamics of structural change: agricultural employment grows slowly or declines in initially agricultural places, while manufacturing employment grows fastest in those same places.

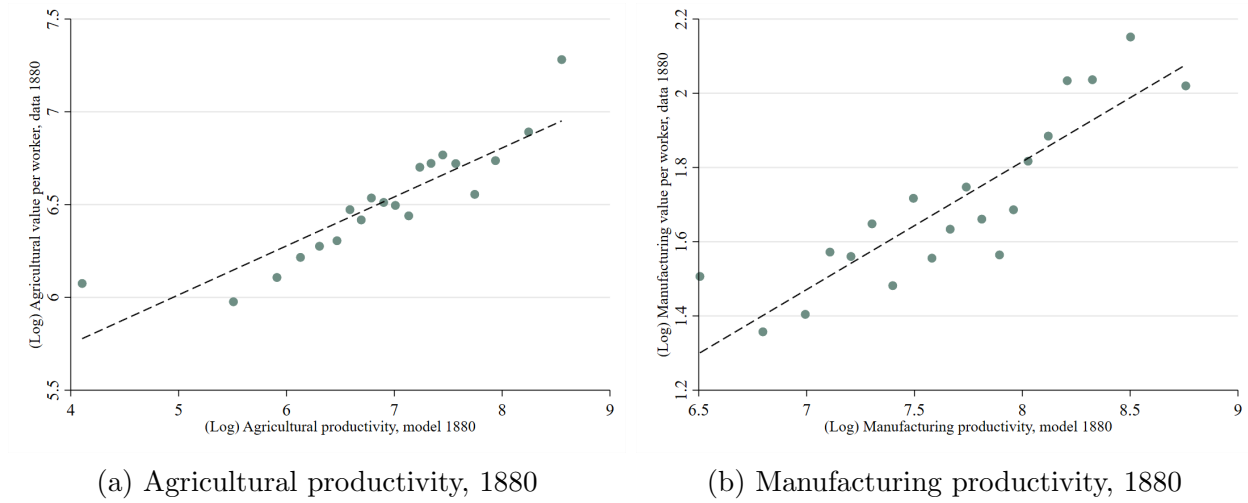
Figure A4.6.2: Geographical distribution of amenities, 1880



Notes: Recovered sectoral amenities across commuting zones in 1880.

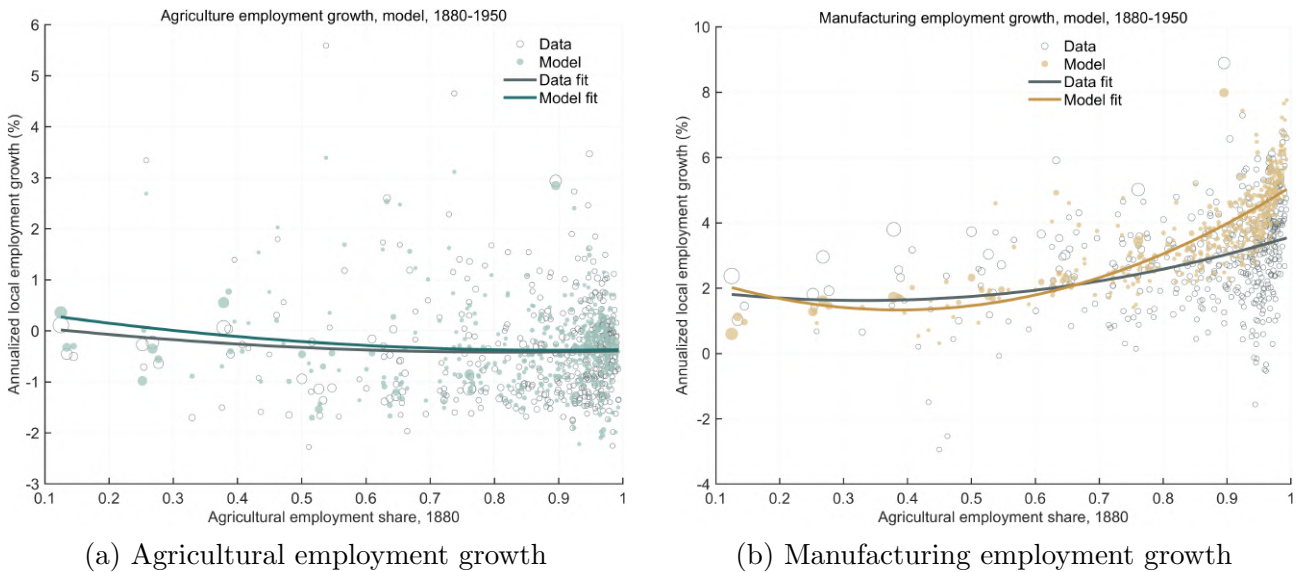
Kin structure and within-region structural change. Figure A4.6.5 provides two additional validations of the kin block. Panel (a) compares the model-implied and observed commuting-zone distribution of uncles and aunts in 1880, an untargeted kinship moment. The close fit indicates that the network law of motion generates a realistic spatial distribution of extended kin. Panel (b) decomposes the 1880–1950 decline in the agricultural share into within-region and across-region components. The model matches the data decomposition: structural change occurred primarily within commuting zones rather than through migration across regions. When kin channels are shut down, the transition becomes larger and even more strongly within-region, confirming that kin access slows sectoral switching for workers who remain in place.

Figure A4.6.3: Recovered productivity versus independent data



Notes: Model-recovered agricultural and manufacturing productivity against agricultural and manufacturing value per worker in the data, by commuting zone.

Figure A4.6.4: Untargeted validation: employment growth, 1880–1950

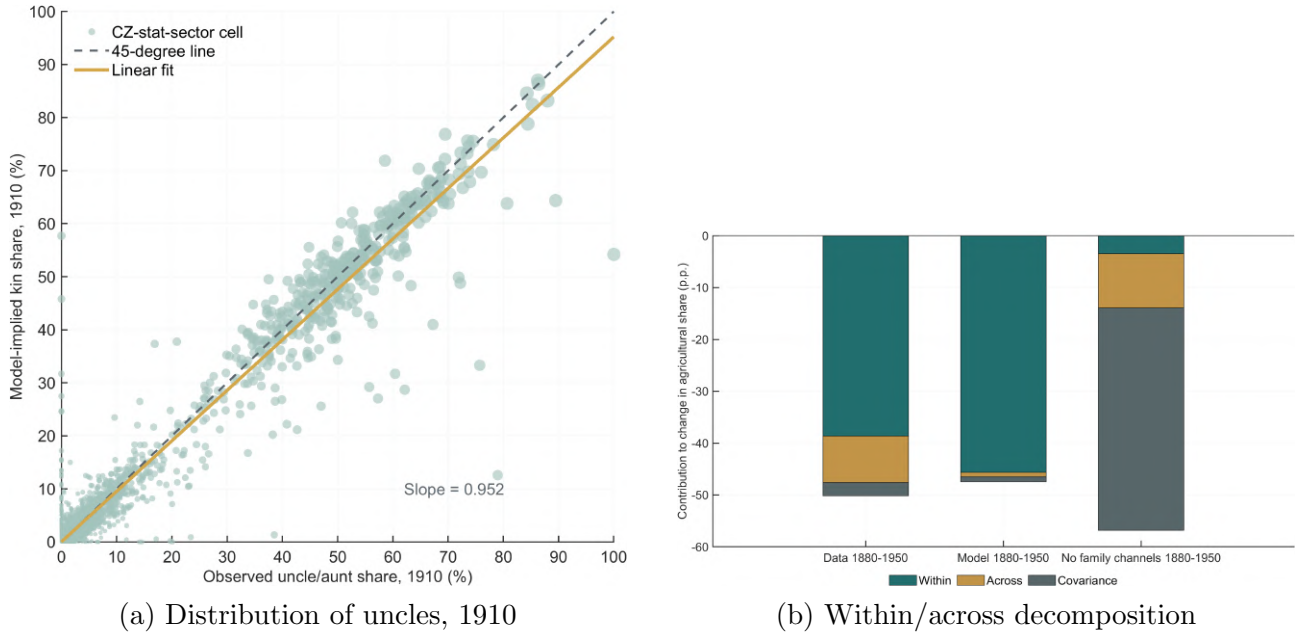


Notes: Annualized local employment growth from 1880 to 1950 against the 1880 agricultural employment share, by commuting zone, in the model and in the data.

A4.7 Representative-kin approximation

The quantitative model does not follow every family tree separately after the initial period. Instead, it groups workers by ancestral origin, sector and location and represents their horizontal kin by the average kin distribution within that group. This greatly reduces the number of family

Figure A4.6.5: Untargeted validation: social ties and within-region reallocation



Notes: Panel (a) reports the model-implied versus observed commuting-zone share of uncles and aunts in 1910. Panel (b) decomposes the 1880–1950 change in the aggregate agricultural share into within-region and across-region components, in the data, the model and a counterfactual without social ties channels.

histories the model must track. This appendix uses the linked genealogical data to assess what is lost through this averaging.

The ancestral type g is not an individual family. It is defined by the commuting zone of the linked 1850 ancestor, so many distinct genealogical families can share the same g . I use f below to index these individual families.

The approximation. Consider adults of ancestral type g who live in m and work in j . The model summarizes their horizontal kin by

$$\Phi_{g,t}(m, j; n, k) = r_g \lambda_{g,t}(n, k \mid m, j), \quad \sum_{n,k} \lambda_{g,t}(n, k \mid m, j) = 1, \quad (\text{A4.7.1})$$

where r_g is total horizontal-kin mass for ancestral type g and $\lambda_{g,t}(n, k \mid m, j)$ is the share of that kin living in location n and working in sector k . Thus, $\Phi_{g,t}$ is the average kin distribution among workers with the same ancestral origin, sector and location.

Descendants from parental state (m, j) choose sectors and locations according to $P_t^O(k \mid m, j, g)$ and $P_{t+1}^N(n \mid m, k, g)$. Adults reaching the same new state (n, k) can come from different

parental states. Their parental composition is

$$\Pr_t(m, j \mid n, k, g) = \frac{L_{gmt}^j P_t^O(k \mid m, j, g) P_{t+1}^N(n \mid m, k, g)}{\sum_{\tilde{m}, \tilde{j}} L_{g\tilde{m}t}^{\tilde{j}} P_t^O(k \mid \tilde{m}, \tilde{j}, g) P_{t+1}^N(n \mid \tilde{m}, k, g)},$$

which gives the law of motion

$$\lambda_{g,t+1}(n', k' \mid n, k) = \sum_{m,j} \Pr_t(m, j \mid n, k, g) P_t^O(k' \mid m, j, g) P_{t+1}^N(n' \mid m, k', g). \quad (\text{A4.7.2})$$

The model therefore represents workers with the same ancestral origin, sector and location by their average kin distribution. This averaging can affect the model in two ways. First, choice probabilities are nonlinear in kin connections, so choices evaluated at the average kin distribution need not equal average choices across individual families. Second, the law of motion follows kin links, so averaging can matter if families that contribute more kin also have systematically different transition rates. I examine both margins below.

Family-level heterogeneity. I first measure how much transition behavior differs across families within the groups used by the model. Let $c = (g, m, j, t)$ denote such a group and let $\hat{p}_{fc}$ be the fraction of observed descendants from family f who enter manufacturing. Because family-level transition rates are noisy when only a few descendants are observed, I subtract the variation expected from finite family size:

$$V_c^{\text{ex}} = \text{Var}_f(\hat{p}_{fc}) - \mathbb{E}_f \left[\frac{\bar{p}_c(1 - \bar{p}_c)}{N_{fc}} \right],$$

where N_{fc} is the number of observed descendants and $\bar{p}_c$ is the group-level transition rate. I also report

$$\text{RMSE}_c^{\text{net}} = \left[\max \left\{ \mathbb{E}_f [(\hat{p}_{fc} - \bar{p}_c)^2] - \mathbb{E}_f \left[\frac{\bar{p}_c(1 - \bar{p}_c)}{N_{fc}} \right], 0 \right\} \right]^{1/2},$$

which measures the typical family-level deviation from the group average after removing variation due to finite family size.

Choices at average versus individual kin distributions. I next test directly whether replacing individual kin distributions by their group average changes sector and location choices. I hold wages, prices, switching costs, migration costs, parental connections and continuation values fixed at their 1880 equilibrium values. For each worker i , let Φ_i denote the observed

horizontal-kin distribution and let $\bar{\Phi}_c$ denote its mean within $c = (g, m, j)$. I compare

$$J_c^O = \frac{1}{N_c} \sum_{i \in c} P^O(M | \Phi_i) - P^O(M | \bar{\Phi}_c), \quad J_{cn}^N = \frac{1}{N_c} \sum_{i \in c} P^N(n | \Phi_i) - P^N(n | \bar{\Phi}_c). \quad (\text{A4.7.3})$$

The first term averages choice probabilities evaluated separately at each worker's observed kin distribution. The second evaluates the same choice problem once at the group-average kin distribution. Their difference therefore measures the effect of averaging kin connections while holding all other equilibrium objects fixed.

Individual kin distributions are constructed from recorded aunts and uncles working in agriculture or manufacturing in modeled CZs. Within each ancestral group, I rescale observed kin by a common factor so that average total horizontal-kin mass equals the calibrated r_g , preserving differences in kin composition and size across individuals. Workers with no recorded eligible kin remain in the sample with zero kin connections whenever their ancestral group contains positive observed kin. I retain groups with at least ten linked children and five parent households. The resulting sample contains 21,808 children in 412 parental-state groups; six children belonging to ancestral groups with no recorded eligible kin are excluded.

For sector choice, I recover the individual-specific index from the stored equilibrium probability p_c^0 :

$$\text{logit } P_i^O(M) = \text{logit } p_c^0 + \frac{b_M(\Phi_i^M - \Phi_c^{0,M}) - b_A(\Phi_i^A - \Phi_c^{0,A})}{\sigma_O},$$

where Φ_i^k is horizontal kin in sector k at the child's birth CZ, $b_A = 0.075213$, $b_M = 0.0826542$ and $\sigma_O = 0.6$. For location choice, letting $H_i(n) = \sum_k \Phi_i^k(n)$, I compute

$$P_i^N(n | m, k, g) \propto P_c^{N,0}(n | k) \exp \left\{ \frac{b_H[H_i(n) - H_c^0(n)]}{\sigma_N} \right\}, \quad b_H = 0.20976, \quad \sigma_N = 0.51.$$

Evaluating the same expressions at $\bar{\Phi}_c$ gives the group-average comparison. Migration probabilities are averaged across the two chosen sectors using the stored sector-choice probabilities, so this exercise isolates averaging in the location decision.

Table A4.7.1: Choice probabilities using individual and average kin distributions

Choice	Average kin (%)	Individual kin (%)	Signed gap (p.p.)	Mean absolute gap (p.p.)
Manufacturing	25.39	25.43	0.046	0.048
Stay in birth CZ	83.73	83.14	-0.586	0.586

Notes: "Individual kin" reports the average probability obtained by evaluating the choice separately for each observed kin distribution. "Average kin" evaluates the same choice at the group-average kin distribution. Signed gap is individual kin minus average kin. Mean absolute gap is the child-weighted mean absolute difference across groups. Migration probabilities are additionally weighted by stored sector-choice probabilities. All other choice inputs are held fixed. The sample contains 21,808 children in 412 parental-state groups for both decisions.

Kin-weighted transitions. The second question is whether families that contribute more kin also have systematically different transition rates. Let K_{fc} denote the horizontal-kin mass contributed by family f and p_{fc} its transition rate. The transition relevant for the law of motion is $\mathbb{E}_f[K_{fc}p_{fc} \mid c]$. Replacing family-specific kin mass and transition rates by their group averages instead gives

$$\mathbb{E}_f[K_{fc} \mid c] \mathbb{E}_f[p_{fc} \mid c].$$

Their difference is

$$\text{Cov}_c(K_f, p_f) = \mathbb{E}_f[K_{fc}p_{fc} \mid c] - \mathbb{E}_f[K_{fc} \mid c]\mathbb{E}_f[p_{fc} \mid c].$$

I report this difference relative to the exact kin-weighted transition,

$$C_c = \frac{\text{Cov}_c(K_f, p_f)}{\mathbb{E}_f[K_{fc}p_{fc} \mid c]}.$$

A value close to zero means that averaging across families has little effect on the kin flow carried into the next generation. I construct the same statistic for the joint sector-location transition, which is closest to the object propagated in equation (A4.7.2).

Table A4.7.2: Diagnostics for the Representative-Family Approximation

Year	Cells	Share pos. excess	Median excess	Median RMSE	W. median RMSE	W. median rel. cov.	W. median joint resid.	Median occ.– mig. cov.
1880	434	0.923	0.018	0.122	0.141	0.039	0.020	0.000
1910	438	0.961	0.020	0.138	0.141	0.047	0.016	0.000
1930	407	0.926	0.012	0.104	0.107	0.031	0.017	0.000

Notes: The table reports diagnostics for the representative-family approximation using all (g, n, k, t) cells. “Share pos. excess” is the share of cells in which family-level sector transition rates exhibit excess variance relative to binomial sampling noise. RMSE is the de-noised root mean squared deviation between family-level transition rates and the cell-level representative-family transition rate. Weighted statistics use family-capital cell mass as weights. “W. median rel. cov.” is the mass-weighted median of the relative covariance between family-specific sector probabilities and family-capital stocks. “W. median joint resid.” compares the joint transition object $E[P_i^O P_i^N K_i]$ to the representative-family counterpart. “Median occ.–mig. cov.” is the within-cell covariance between sector and migration probabilities.

Results. The data show substantial family-level heterogeneity: more than 90% of groups have positive excess variance and the adjusted RMSE is roughly 10–14 percentage points. Despite this heterogeneity, evaluating choices at the group-average kin distribution produces only small aggregate differences. The manufacturing probability changes by 0.046 percentage points and the probability of remaining in the birth CZ by 0.586 points.

The kin-weighted transition check gives a similar conclusion. The median discrepancy is approximately 3–5% for sector transitions and about 2% for joint sector-location transitions. The

model therefore averages over meaningful differences across individual families, but the two margins through which this averaging affects aggregate choices and kin flows are considerably smaller.

A5 Quantitative analysis

A5.1 Solution algorithm

I solve the model by separating the within-period temporary equilibrium from the intertemporal sequential equilibrium, following the logic of [Caliendo et al. \(2019\)](#) and [Eckert and Kleineberg \(2024\)](#). Conditional on a guess for labor, kin connections and continuation values, wages and prices can be solved independently at each date. These market outcomes determine sector and location choices. The resulting choices then update the labor and kin distributions and value functions. I iterate on the full path until the market equilibrium, choices, laws of motion and value functions are mutually consistent.

Kin decomposition. All counterfactuals use the same recovered non-kin fundamentals and structural parameters and start from the same observed initial labor allocation. For the baseline inheritance and propagation exercises, parental connections remain active throughout. I consider three treatments of horizontal kin. In the no-horizontal-ties economy, I set the horizontal terms in equations (8) and (12) to zero. In the constant-1880 economy, I hold the observed 1880 horizontal-kin distribution constant at all subsequent dates. In the endogenous economy, the horizontal-kin distribution evolves according to equations (15)–(17). Within each kin regime, I compare the policy economy with the same economy without the policy. A complementary exercise in Appendix below also holds investigates the strength of parents.

For any equilibrium outcome Y_t , the effect of the initial horizontal-kin distribution and the additional effect of endogenous propagation can be decomposed as

$$\mathcal{I}_t = Y_t^{\text{constant}} - Y_t^{\text{none}}, \quad (\text{A5.1.1})$$

$$\mathcal{P}_t = Y_t^{\text{endo}} - Y_t^{\text{constant}}. \quad (\text{A5.1.2})$$

Temporary equilibrium. Fix a date t , a trial labor distribution L_t and the date- t fundamentals. First aggregate employment over ancestral types,

$$L_{nt}^k = \sum_g L_{gnt}^k, \quad L_{nt} = \sum_{k \in \{A, M\}} L_{nt}^k.$$

Conditional on these quantities, the temporary equilibrium consists of wages, prices, trade shares and expenditures $\{w_t, p_t, \pi_t, X_t\}$. For a trial wage vector, equations (18) and (19) determine trade shares and sectoral prices. Equations (5) and (22) then determine sectoral expenditures. Wages are updated until the production-market conditions in equation (21) hold in every sector and location. I normalize one nominal wage because only relative nominal prices are identified.

Algorithm 1 Temporary equilibrium at date t

Require: Trial labor distribution L_t , date- t fundamentals and structural parameters.

- 1: Aggregate $L_{nt}^k = \sum_g L_{gnt}^k$ and $L_{nt} = \sum_k L_{nt}^k$.
 - 2: Initialize wages $w_t^{(0)}$.
 - 3: Set the market-clearing residual above tolerance.
 - 4: **while** the market-clearing residual is above tolerance **do**
 - 5: Compute trade shares π_t from equation (18).
 - 6: Compute sectoral price indices p_t from equation (19).
 - 7: Compute PIGL expenditure shares $\delta_k(w_{nt}^j, p_{nt})$ from equation (5).
 - 8: Aggregate sectoral expenditures X_t from equation (22).
 - 9: Evaluate excess demand using equation (21).
 - 10: Update wages.
 - 11: **end while**
 - 12: **return** $\{w_t, p_t, \pi_t, X_t\}$.
-

This block is static conditional on L_t . Horizontal kin do not enter equations (18)–(22) directly. They affect the temporary equilibrium only through the labor distribution generated by earlier sector and location choices.

Sequential equilibrium. The dynamic solution requires the labor and continuation-value paths used to make choices to agree with the paths generated by those choices. Let $\mathcal{T} = \{1880, 1910, 1930, 1950, 1970\}$ denote the historical solution dates and let $\mathcal{T}^E$ additionally include the terminal extension used to evaluate continuation values. I iterate on $\{L_t, V_t\}_{t \in \mathcal{T}^E}$, holding the observed initial state fixed. The trial labor path determines wages and prices at each date, while the kin distribution Φ_t is generated forward from the initial distribution. The terminal extension uses the same value function as in 1950.

Each iteration has four steps. First, I solve wages and prices at every date given the trial labor path. Second, I move forward through time: children choose sectors, then locations as adults and these choices determine the next period's labor and kin distributions. Third, I move backward through time and recompute opportunity and continuation values using the paths generated in the forward step. Finally, I update the trial labor and continuation-value paths

and repeat until they agree with the paths implied by the model. Sector and location choices can be solved sequentially because the kin connections relevant for each transition are inherited from the previous generation.

Algorithm 2 Sequential equilibrium and endogenous kin propagation

Require: Initial state (L_{t_0}, Φ_{t_0}) , path of fundamentals, terminal extension and a kin regime.

- 1: Initialize trial paths $\{L_t^{(0)}, V_t^{(0)}\}_{t \in \mathcal{T}^E}$.
- 2: **while** the outer fixed-point residual is above tolerance **do**
 - A. *Temporary equilibria*
 - 3: **for** each historical date $t \in \mathcal{T}$ **do**
 - 4: Solve $\{w_t^{(s)}, p_t^{(s)}, \pi_t^{(s)}, X_t^{(s)}\}$ given $L_t^{(s)}$ with Algorithm 1.
 - 5: **end for**
 - B. *Forward pass*
 - 6: Set $(\tilde{L}_{t_0}, \Phi_{t_0})$ to the observed initial state.
 - 7: **for** each historical transition $t \rightarrow t + 1$ **do**
 - 8: Construct sector-specific kin connections from Φ_t , form sector values using $V_{t+1}^{(s)}$ and compute P_t^O from equation (9).
 - 9: Compute children's opportunity values O_t using the kin distribution from and the updated continuation value at $t + 1$, with the terminal extension used after the final date.
 - 10: Construct destination-specific kin amenities and form destination values using the trial market outcomes at $t + 1$. Compute P_{t+1}^N from equation (13).
 - 11: Update $\tilde{L}_{t+1}$ from equation (20).
 - 12: **if** horizontal kin are endogenous **then**
 - 13: Update Φ_{t+1} using equations (15)–(17).
 - 14: **else if** the 1880 horizontal-kin distribution is constant **then**
 - 15: Set $\Phi_{t+1} = \Phi_{1880}$.
 - 16: **else**
 - 17: Set the horizontal component of Φ_{t+1} to zero.
 - 18: **end if**
 - 19: **end for**

Algorithm 2 Sequential equilibrium and endogenous kin propagation (continued)

C. Backward pass

- 20: **for** each historical date $t \in \mathcal{T}$, from the last to the first **do**
 21: Compute updated continuation values $\tilde{V}_t$ from equation (14), using the wages and prices from block A and the kin distributions from block B.
 22: **end for**

D. Update

- 23: Compute the maximum change between $\{\tilde{L}_t, \tilde{V}_t\}$ and the current trial paths together with the maximum market-clearing residual.
 24: Update the labor path by damping: $L_t^{(s+1)} = (1 - \omega)L_t^{(s)} + \omega\tilde{L}_t$.
 25: Update continuation values analogously: $V_t^{(s+1)} = (1 - \omega)V_t^{(s)} + \omega\tilde{V}_t$, $0 < \omega \leq 1$,
 leaving the initial state and terminal extension fixed.
 26: **end while**
 27: **return** the converged paths $\{L_t, \Phi_t, V_t, O_t, P_t^O, P_t^N, w_t, p_t, \pi_t, X_t\}$.
-

At convergence, wages and prices clear markets, workers' sector and location choices are optimal and the labor and continuation-value paths generated by those choices coincide with the paths used to solve the model.

Moving voucher construction. The grant in equation (A5.5.3) enters only the payoff from eligible origin–destination moves and the common lump-sum tax is included in household income. The voucher affects equilibrium wages and prices indirectly because it changes location choices and therefore subsequent labor allocations.

Calibrating the grant is an outer root problem around Algorithm 2. For a candidate G , I solve the full sequential equilibrium, compute realized take-up $F(G)$ and expenditure $GF(G)$ and update the grant until expenditure equals the target in equation (A5.5.4). Because take-up is endogenous, the equilibrium is re-solved after every change in G . When the benchmark policy is compared across kin regimes, the grant is recalibrated separately within each regime so that realized fiscal expenditure is held constant.

Implementation of the additional counterfactuals. The additional exercises use the same solution algorithm and change one element at a time. To measure the effect of kin connections created by the relocation policy, I keep the policy-induced 1880 sector and location allocation unchanged but, from 1910 onward, give each cohort the parental and horizontal-kin connections it would have had without the policy. Wages, prices and choices are then allowed to adjust normally. The difference from the full policy economy therefore captures the effects of the new kin connections created by the initial relocation and prevents the comparison to a counterfactual world where the ties would be constant.

The destination-targeting exercises instead change which origin–destination moves are eligible for the grant. When comparing high- and low-kin destinations at the same initial migration response, I choose the grant in each case to generate the same increase in 1880 movers rather than the same expenditure. Finally, the channel decompositions change one component of kin connections at a time and re-solve the model.

A5.2 Additional aggregate effects of endogenous kin adjustment

Section 6.1 shows that kin connections account for sizable parts of the sectoral and spatial barriers faced by individual workers. This appendix reports additional aggregate outcomes from allowing horizontal kin connections to evolve with sector and location choices rather than remaining constant at their 1880 distribution.

A5.3 The initial horizontal-kin distribution

This appendix reports additional results underlying Section 6.1. The main text focuses on how the initial horizontal-kin distribution affects the manufacturing share. I first report corresponding effects on migration, the wage bill and geographic dispersion. I then show how these effects evolve over time and separate the role of kin quantity from the role of kin geography.

Additional outcomes. Table A5.3.1 reports the 1950 effects of the initial horizontal-kin distribution on manufacturing, migration, the wage bill and cross-CZ dispersion in manufacturing shares. As in the main text, I separate agricultural, manufacturing and local horizontal-kin connections by introducing the three channels in every possible order and averaging their marginal contributions.

Table A5.3.1: Channels of the initial horizontal-kin distribution

	Mfg. share (p.p.)	Migration (p.p.)	Wage bill (%)	Mfg. dispersion (p.p.)
Agricultural connections	-1.200	-0.344	-1.250	0.549
Manufacturing connections	0.142	0.024	0.156	0.054
Local connections	-0.253	-1.704	-0.502	0.154
Total initial-distribution effect	-1.311	-2.024	-1.596	0.757

Notes: Exact three-factor Shapley decomposition of the 1950 effect of holding the observed 1880 horizontal-kin distribution constant relative to an economy without horizontal kin. Parental connections and all non-kin fundamentals remain unchanged. Components add to totals before rounding.

The results show a clear distinction between the sectoral and spatial channels. Agricultural connections account for most of the decline in manufacturing and the wage bill, while local

connections account for most of the decline in migration. The initial kin distribution also increases cross-CZ dispersion in manufacturing shares, consistent with greater persistence in the sectoral and spatial allocation inherited from 1880.

Initial horizontal-kin effect over time. Table A5.3.2 traces the same decomposition across model dates for the manufacturing share and the wage bill. Parents evolve in both economies, so the comparison continues to isolate the effect of the horizontal-kin distribution inherited in 1880.

Table A5.3.2: Initial horizontal-kin distribution: channels

	1910	1930	1950	1970
<i>Panel A. Manufacturing share (p.p.)</i>				
Agricultural connections	−0.69	−1.19	−1.20	−1.20
Manufacturing connections	0.08	0.10	0.14	0.08
Local connections	0.37	0.32	−0.25	−1.16
Total	−0.25	−0.77	−1.31	−2.28
<i>Panel B. Wage bill (%)</i>				
Agricultural connections	−0.98	−1.53	−1.25	−1.15
Manufacturing connections	0.12	0.16	0.16	0.09
Local connections	0.34	0.02	−0.50	−1.13
Total	−0.52	−1.35	−1.60	−2.19

Notes: Difference between an economy with horizontal kin held at their observed 1880 distribution and an economy without horizontal kin. Parents evolve in both economies. The three rows give the Shapley contributions of agricultural, manufacturing and local connections. The wage bill is measured relative to the economy without horizontal kin.

The initial effect becomes more negative over time. Agricultural connections account for most of the early decline in manufacturing and the wage bill. The contribution of local connections is initially positive but becomes increasingly negative, accounting for a large share of the additional decline by 1970. Manufacturing connections partly offset these effects throughout.

Kin quantity and kin geography. The comparison with an economy without horizontal kin combines two features of the observed 1880 distribution: workers have horizontal kin and those relatives are unevenly distributed across space. To separate these margins, I construct an intermediate economy that preserves total horizontal-kin mass within each ancestral type but distributes relatives across CZs in proportion to population. Table A5.3.3 reports the results. The geography of kin accounts for nearly the entire manufacturing effect. When the same horizontal-kin mass is instead distributed across CZs in proportion to population, the manufacturing share falls by only 0.030 percentage points. The additional effect of the observed geography is −1.281 points. The pattern is even stronger for migration.

Table A5.3.3: Quantity and geography of the initial horizontal-kin distribution, 1950

	Mfg. share (p.p.)	Migration (p.p.)	Wage bill (%)	Mfg. dispersion (p.p.)
Observed initial horizontal kin	-1.311	-2.024	-1.596	0.757
Kin quantity, population-proportional geography	-0.030	0.064	0.010	-0.238
Observed kin geography	-1.281	-2.089	-1.606	0.995

Notes: The first row compares the observed 1880 horizontal-kin distribution with an economy without horizontal kin. The second preserves the same total horizontal-kin mass but distributes relatives across CZs in proportion to population. The final row is the difference between the observed distribution and this population-proportional benchmark. Parental connections and all non-kin fundamentals remain unchanged.

Sectoral change within CZs and employment reallocation across CZs. Finally, I ask whether the lower national manufacturing share comes from lower manufacturing shares within CZs or from a shift of employment toward CZs with lower manufacturing shares. For counterfactual c ,

$$\Delta S_c^M = \underbrace{\sum_n e_{n0} \Delta s_{nc}^M}_{\text{within CZs}} + \underbrace{\sum_n \Delta e_{nc} s_{n0}^M}_{\text{across CZs}} + \underbrace{\sum_n \Delta e_{nc} \Delta s_{nc}^M}_{\text{interaction}}, \quad (\text{A5.3.1})$$

where e_{n0} denotes the benchmark employment share of CZ n and s_{n0}^M its manufacturing share.

Table A5.3.4: Within- and across-CZ effects of initial horizontal kin, 1950

	Within CZs	Across CZs	Interaction	Total
Initial horizontal kin	-0.688	-0.603	-0.020	-1.311

Notes: Decomposition of the 1950 manufacturing-share effect of the initial horizontal-kin distribution. “Within CZs” captures changes in local manufacturing shares holding the spatial distribution of employment at its benchmark value. “Across CZs” captures changes in the spatial distribution of employment holding local manufacturing shares at their benchmark values. The interaction captures the joint movement of the two margins. Entries are percentage points and add to the total before rounding.

The within- and across-CZ components are similar in magnitude. The initial horizontal-kin distribution therefore lowers the national manufacturing share through both lower manufacturing shares within places and a spatial distribution of employment tilted toward less manufacturing-intensive places.

A5.4 Endogenous propagation of horizontal kin

This appendix reports additional results underlying Section 6.2. Throughout the main comparison, endogenous propagation compares the full model, in which horizontal kin evolve with workers’ sector and location choices, with an otherwise identical economy in which horizontal kin remain constant at their observed 1880 distribution. Parental connections and all non-kin

fundamentals evolve identically in both economies. I then report a complementary exercise in which parental sector connections are also held at their 1880 distribution.

Construction of the main-text statistics. Let Y_t^{full} denote an outcome in the full model and Y_t^{constant} the same outcome when horizontal kin remain at their observed 1880 sectoral and spatial distribution. Because the two economies start from the same 1880 allocation, the share of the 1880–1970 change in outcome Y explained by evolving kin, reported in Panel C of Table 4, is

$$\mathcal{S}_Y = 100 \frac{Y_{1970}^{\text{full}} - Y_{1970}^{\text{constant}}}{Y_{1970}^{\text{full}} - Y_{1880}^{\text{full}}}. \quad (\text{A5.4.1})$$

The same calculation is applied to the national manufacturing share, the manufacturing share of the initially least industrialized third of CZs, the top–bottom manufacturing-share gap and the dispersion of CZ manufacturing shares. A positive share therefore means that allowing kin to evolve moves the outcome in the same direction as its full-model change between 1880 and 1970.

The bottom and top thirds are defined once using CZ manufacturing shares in 1880 and are held constant thereafter. A group’s manufacturing share at date t is total manufacturing employment in that group divided by its total employment. The South is the Census South region. The population share living in manufacturing-majority CZs is

$$\mathcal{M}_t = \sum_n e_{nt} \mathbf{1}\left\{s_{nt}^M > \frac{1}{2}\right\}, \quad (\text{A5.4.2})$$

where e_{nt} is CZ n ’s share of national employment and s_{nt}^M its manufacturing share. I measure cross-CZ dispersion by the employment-weighted standard deviation

$$\text{GeoSD}_t = \left[\sum_n e_{nt} (s_{nt}^M - S_t^M)^2 \right]^{1/2}, \quad (\text{A5.4.3})$$

where S_t^M is the national manufacturing share. The top–bottom gap is the manufacturing share of CZs in the top 1880 third minus that of CZs in the bottom third. For the timing statistic, I linearly interpolate the full-model manufacturing share between model dates and record the first year in which it reaches the constant-1880 economy’s 1970 manufacturing share.

Additional aggregate outcomes. The main text focuses on the manufacturing share and its spatial incidence. Table A5.4.1 reports corresponding effects on the wage bill, agricultural employment, migration and the dispersion of manufacturing shares across CZs.

Table A5.4.1: Additional consequences of changing horizontal kin connections

	Evolving vs. held at 1880
Wage bill in 1970 (%)	+1.78
Cumulative wage bill, 1910–1970 (%)	+0.86
Agricultural employment in 1970 (%)	−9.2
Migration rate in 1970 (p.p.)	+1.13
SD of CZ manufacturing shares (p.p.)	−1.23
CZs with a manufacturing majority	+13

Notes: Effects compare the full model with an otherwise identical economy in which horizontal kin remain at their observed 1880 sectoral and spatial distribution. The wage bill, cumulative wage bill and agricultural employment are percentage changes relative to the constant-horizontal-kin economy. Migration and manufacturing dispersion are percentage-point changes.

Detailed propagation channels. Table A5.4.2 reports the detailed decomposition underlying Panel 8a. Spatial propagation is separated into persistence at the origin inherited from the preceding generation and connections to destinations. Sectoral propagation is separated analogously into same-sector persistence and connections to the alternative sector.

Table A5.4.2: Channels of endogenous horizontal-kin propagation

	1950	1970
<i>Spatial propagation</i>		
Origin persistence	−0.161	0.608
Destination pull	0.172	0.305
<i>Spatial total</i>	0.009	0.911
<i>Sectoral propagation</i>		
Sector persistence	0.502	0.738
Cross-sector pull	0.214	0.402
<i>Sectoral total</i>	0.718	1.142
Total propagation	0.727	2.053

Notes: Contributions to the national manufacturing share in percentage points. Components are obtained from the hierarchical Shapley decomposition used in Figure 8a. Components may not add exactly because of rounding.

Spatial propagation is close to zero in 1950 because origin persistence and destination pull offset one another, while sectoral propagation is already positive. By 1970, the two channels contribute 0.911 and 1.142 percentage points to the manufacturing share.

Within-CZ adjustment and propagation across generations. Table A5.4.3 reports two complementary accounting exercises. Panel A separates the manufacturing effect into changes

in sectoral composition within CZs and changes in the spatial distribution of employment. Panel B varies how long horizontal kin are allowed to evolve before their distribution is held constant.

Table A5.4.3: Where and when horizontal-kin propagation operates

	1910	1930	1950	1970
<i>Panel A. Manufacturing effect within and across CZs</i>				
Within CZs (p.p.)	-0.3	-0.2	0.5	1.5
Across CZs (p.p.)	0.0	0.1	0.2	0.5
<i>Panel B. 1970 manufacturing effect by propagation horizon</i>				
Horizontal kin evolve through 1910		-0.132 p.p.		
Horizontal kin evolve through 1930		1.522 p.p.		
Horizontal kin evolve through 1950		1.829 p.p.		
Horizontal kin evolve throughout		2.053 p.p.		

Notes: Panel A decomposes the difference between the full model and the economy with horizontal kin held at their 1880 distribution into changes in CZ manufacturing shares and changes in the spatial distribution of employment. Panel B reports the 1970 manufacturing-share effect when horizontal kin are allowed to evolve only through the indicated date and are then held constant. The comparison economy in Panel B holds horizontal kin at their 1880 distribution throughout.

By 1970, approximately three quarters of the manufacturing effect operates within CZs. Most of the long-run gain emerges only after additional generations of sector switching and migration.

Reorientation of horizontal kin. Table A5.4.4 reports the changes in horizontal-kin composition that underlie propagation.

Table A5.4.4: Reorientation of horizontal kin connections

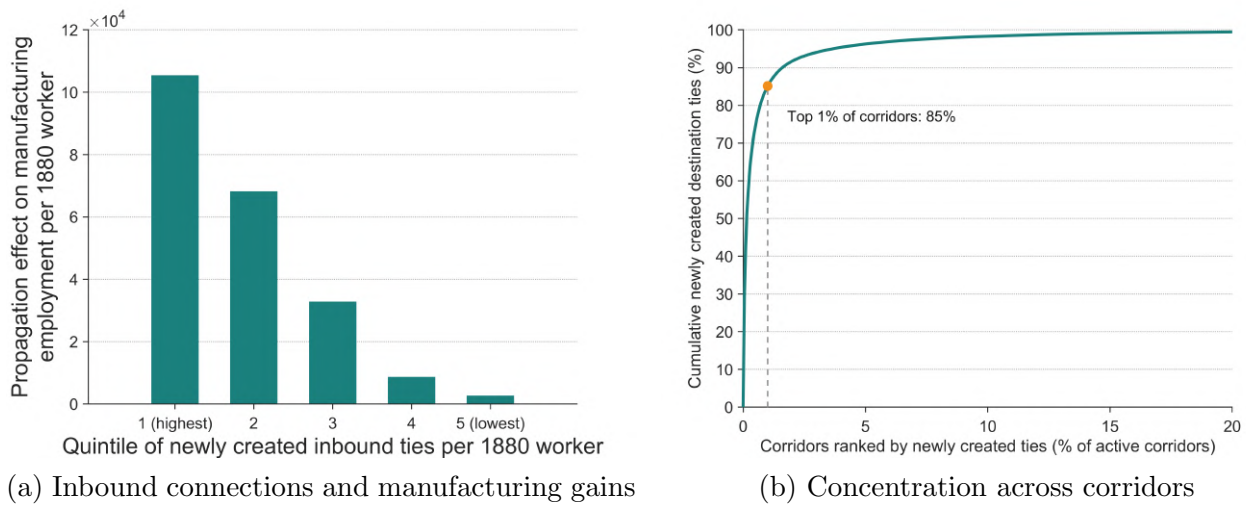
	1880	1910	1930	1950	1970
<i>Panel A. Sectoral orientation</i>					
Agriculture persistence	0.452	0.528	0.326	0.276	0.223
Agriculture-to-manufacturing	0.044	0.107	0.159	0.213	0.207
Manufacturing-to-agriculture	0.261	0.132	0.110	0.076	0.053
Manufacturing persistence	0.116	0.234	0.321	0.423	0.463
<i>Panel B. Sector supplying new destination connections</i>					
Agriculture (%)	—	16.4	18.4	13.9	13.4
Manufacturing (%)	—	83.6	81.6	86.1	86.6

Notes: Panel A reports mass-weighted horizontal-kin connections by the sectors of successive cohorts. Persistence denotes connections remaining within the same sector, while the two cross-sector rows denote connections spanning sectors. Panel B splits newly created destination connections by the sector of the relative supplying the connection.

Manufacturing persistence roughly quadruples between 1880 and 1970 while agriculture persistence halves. Most new destination connections are supplied by manufacturing relatives, so sectoral and spatial propagation increasingly reinforce one another.

Geography of destination connections. Figure A5.4.1 provides two additional views of spatial propagation. Panel A5.4.1a relates newly created inbound connections to the manufacturing gains generated by propagation, while Panel A5.4.1b shows how concentrated new destination connections are across origin–destination corridors.

Figure A5.4.1: Destination connections and the geography of propagation



Notes: Panel (a) ranks CZs by newly created inbound horizontal-kin connections per 1880 worker and groups them into population-weighted quintiles. Bars report the increase in manufacturing employment per 1880 worker generated by endogenous propagation. Panel (b) plots the cumulative share of gross newly created destination connections against ordered origin–destination pairs ranked from largest to smallest. Only pairs with a positive change are included.

New destination connections are concentrated along particular migration corridors, and CZs receiving more of them experience larger manufacturing gains. Spatial propagation therefore reflects a structured reorientation of kin geography rather than a uniform increase in geographic dispersion.

Allowing parental connections to evolve. The main propagation exercise keeps parental connections endogenous in both economies and therefore isolates the evolution of horizontal kin. As a complementary exercise, I also compare the full model with an economy in which both margins remain at their 1880 pattern. For horizontal kin, I hold their sectoral and spatial distribution constant as above. For parents, I hold each dynasty's distribution of parental sectors at its 1880 mix in every later cohort. The parent's own-sector connection remains present; what is removed is the change in which parental sectors later dynasties inherit. All non-kin fundamentals evolve exactly as in the full model.

I decompose the resulting full-minus-constant effect hierarchically. Let Y_t^{pr} , with $p, r \in \{0, 1\}$, denote the equilibrium outcome when parental connections (p) and horizontal-kin connections (r) are either allowed to evolve (1) or held at their 1880 pattern (0). The parental and horizontal-kin contributions are

$$\Delta_t^P = \frac{1}{2} [(Y_t^{11} - Y_t^{01}) + (Y_t^{10} - Y_t^{00})], \quad (\text{A5.4.4})$$

$$\Delta_t^R = \frac{1}{2} [(Y_t^{11} - Y_t^{10}) + (Y_t^{01} - Y_t^{00})]. \quad (\text{A5.4.5})$$

I then apply the same hierarchical Shapley decomposition as in Table A5.4.2 to Δ_t^R , separating origin persistence, destination pull, sector persistence and cross-sector pull and averaging those contributions over the parental margin. The components therefore add to the full-minus-constant effect before rounding.

Table A5.4.5: Evolving kin connections, including parents: channels

	1910	1930	1950	1970
<i>Panel A. Manufacturing share (p.p.)</i>				
Origin persistence	−0.39	−0.39	−0.11	0.57
Destination pull	0.01	0.02	0.14	0.33
Sector persistence	0.06	0.23	0.50	0.77
Cross-sector pull	−0.02	0.11	0.16	0.36
Parents	0.17	2.60	3.84	6.96
Total	−0.17	2.58	4.53	8.99
<i>Panel B. Wage bill (%)</i>				
Origin persistence	−0.51	−0.40	−0.09	0.50
Destination pull	0.01	0.02	0.09	0.25
Sector persistence	0.09	0.29	0.54	0.81
Cross-sector pull	−0.02	0.14	0.18	0.37
Parents	0.25	3.48	4.38	7.51
Total	−0.20	3.53	5.10	9.44

Notes: Difference between the full model and an economy with both parental and horizontal-kin connections held at their 1880 pattern. Hierarchical Shapley decomposition: parents against horizontal kin, with the horizontal-kin component split into origin persistence, destination pull, sector persistence and cross-sector pull and averaged over parents evolving and held. The parental tie rewards only the parent's own sector wherever the child lives, so it acts as sector persistence. The wage bill is measured relative to the constant-1880 economy.

The parent-inclusive exercise uses the same statistics as equations (A5.4.1)–(A5.4.3). Its constant-1880 economy additionally holds parental sector connections at their 1880 mix. This comparison is reported separately because the main text focuses on horizontal kin, which are the margin through which one branch's choices change the connections available to other branches.

Table A5.4.6: Effects of evolving kin connections, including parents, on structural transformation

<i>Panel A. Aggregate structural transformation</i>			
	Full	Constant 1880	Effect
Manufacturing share in 1970 (%)	79.7	70.7	+9.0 p.p.
Year reaching constant-kin 1970 manufacturing share	1946	1970	24 years earlier
<i>Panel B. Spatial incidence</i>			
	Full	Constant 1880	Effect
Population in manufacturing-majority CZs (%)	90.8	82.4	+8.4 p.p.
Manufacturing share in the South (%)	71.3	58.8	+12.6 p.p.
Manufacturing share in bottom-third CZs (%)	62.9	49.7	+13.2 p.p.
<i>Panel C. Contribution to structural change and convergence</i>			
	Full-model change	Share explained by kin	
Manufacturing share (p.p.)	55.1	16%	
Within CZs (p.p.)	48.7	16%	
Across CZs (p.p.)	6.4	21%	
Manufacturing share, bottom-third CZs (p.p.)	60.1	22%	
Top–bottom third manufacturing-share gap (p.p.)	−16.8	43%	
SD of CZ manufacturing shares (p.p.)	−6.5	57%	

Notes: The full model allows kin connections to evolve endogenously with sector and location choices. The constant-kin counterfactual holds both the parental tie, at each dynasty’s 1880 parental sector mix, and the sectoral and spatial distribution of horizontal kin at their observed 1880 pattern, while all non-kin fundamentals evolve as in the full model. Panels A and B report outcomes in the two economies and their difference. Panel C reports the share of the 1880–1970 change in each outcome explained by evolving kin connections, calculated as the difference between the full and constant-kin economies relative to the change in the full model. Thirds of CZs are defined by their manufacturing share in 1880. The within- and across-CZ components come from a shift-share decomposition. The standard deviation of CZ manufacturing shares is weighted by employment. For the timing measure, the full-model path is linearly interpolated between model periods.

Robustness to a smaller kin stock. Table A5.4.7 repeats the decompositions after reducing each dynasty’s stock of non-parental kin, r_g , by 30 percent. The effects operating through these relatives fall roughly proportionally. The 1880 kin distribution lowers the 1970 manufacturing share by 1.44 percentage points rather than 2.27, while allowing these connections to evolve raises it by 1.28 points rather than 2.03. Both effects are therefore about two thirds as large as in the main specification. The ranking of the channels is unchanged: agricultural and local connections account for most of the initial drag, while sector persistence and origin persistence account for most of the subsequent gain. The contribution of parents is almost unchanged because the exercise varies only the stock of non-parental kin. Combining both margins, evolving kin raise the 1970 manufacturing share by 8.18 percentage points, compared

with 8.99 points in the main specification.

Table A5.4.7: Robustness to a smaller non-parental kin stock, 1970

	Manufacturing share (p.p.)		Output (%)	
	Main	$0.7r_g$	Main	$0.7r_g$
<i>Panel A. Initial 1880 kin distribution</i>				
Agricultural connections	−1.19	−0.84	−1.14	−0.80
Manufacturing connections	0.08	0.06	0.10	0.07
Local connections	−1.16	−0.66	−1.13	−0.67
<i>Total</i>	−2.27	−1.44	−2.18	−1.40
<i>Panel B. Evolving kin connections</i>				
Origin persistence	0.57	0.33	0.50	0.28
Destination pull	0.33	0.20	0.25	0.14
Sector persistence	0.77	0.52	0.81	0.54
Cross-sector pull	0.36	0.24	0.37	0.24
Parents	6.96	6.89	7.51	7.38
<i>Total</i>	8.99	8.18	9.44	8.58

Notes: The $0.7r_g$ specification reduces each dynasty’s stock of non-parental kin by 30 percent relative to the main specification. Panel A reports the effect of the 1880 non-parental kin distribution relative to an economy without non-parental kin. Panel B reports the effect of allowing kin connections to evolve rather than keeping their 1880 distribution constant and includes the contribution from evolving parental connections. The parental component is largely unaffected because the exercise changes only r_g . Manufacturing effects are percentage-point changes in the national manufacturing share. Output effects are percent changes.

A5.5 Relocation voucher

Policy eligibility, financing and calibration. The relocation policy is a one-time intervention in 1880. I rank commuting zones by their manufacturing share of employment in 1880 and define the eligible origins $\mathcal{L}$ as the bottom 30 percent. These 137 of the 455 CZs contain 20.9 percent of the 1880 population and 2.25 percent of national manufacturing employment. Eligibility depends only on the observed 1880 allocation and is held constant across all counterfactuals.

The policy pays a common grant G to workers who leave an eligible CZ for a destination outside $\mathcal{L}$. Unless otherwise stated, every destination outside $\mathcal{L}$ is eligible. The budget equals one percent of the 1880 national wage bill,

$$B^* = 0.01 WB_{1880} = \$119.4 \text{ million.} \quad (\text{A5.5.1})$$

It is financed by a uniform lump-sum tax

$$T = \frac{B^*}{\sum_n L_{n,1880}} = \$34.84 \quad (\text{A5.5.2})$$

per person. The tax is applied after wages clear and therefore does not enter firms' labor costs. For a worker in sector j , with origin o and destination d , the consumption component of destination utility is

$$\tilde{v}_{do,1880}^j = v(w_{d,1880}^j - T + G \mathbf{1}\{o \in \mathcal{L}, d \notin \mathcal{L}\}, p_{d,1880}). \quad (\text{A5.5.3})$$

The voucher therefore changes household income at eligible destinations but does not alter the bilateral migration-cost term, productivity or firms' wage costs.

Take-up is endogenous. Let $F(G)$ denote the mass of workers who receive the relocation grant when its value is G . For each kin regime, I choose G such that realized expenditure exhausts the common budget:

$$B_{\text{relocation}}(G) = G F(G) = B^*. \quad (\text{A5.5.4})$$

The grant is solved jointly with the 1880 equilibrium using the outer root described in Appendix A5.1, so changes in migration induced by the policy feed back into take-up and the grant required to satisfy the budget constraint. Comparisons across the benchmark kin regimes therefore hold realized fiscal expenditure constant.

Table A5.5.1: Relocation-voucher calibration and take-up

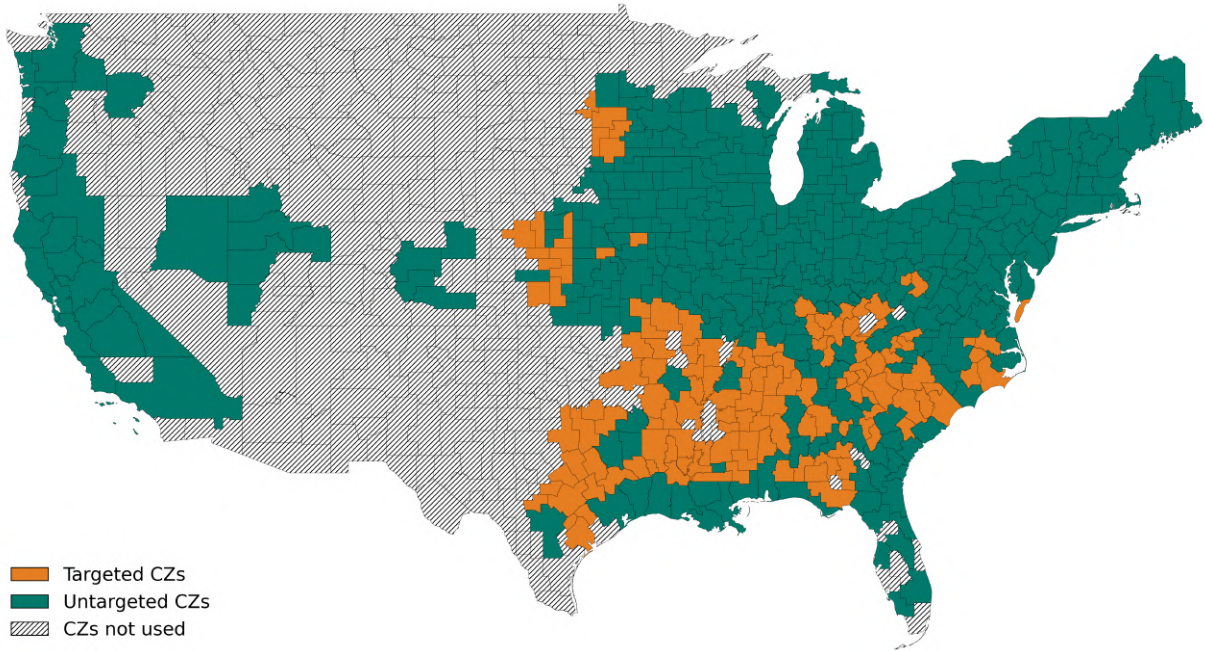
	No horizontal kin	Constant 1880 kin	Endogenous kin
Relocation grant (\$)	707	724	723
Grant recipients (000s)	168.8	164.8	165.1
Recipients / targeted population (%)	23.56	23.01	23.04

Notes: Each counterfactual spends one percent of the 1880 national wage bill. The relocation grant is paid to workers who move from a targeted CZ to a destination outside $\mathcal{L}$. The grant is chosen separately in each kinship regime so that realized expenditure equals the common policy budget. Grant recipients include workers who would have moved even without the policy. Policy-induced movers are therefore defined as the additional 1880 movers generated by the voucher relative to the corresponding no-policy economy and are not the same object as grant recipients.

The policy changes the 1880 allocation, which affects later wages and locations as well as the kin connections inherited by later cohorts. Effects need not therefore have the same sign in every subsequent period.

Figure A5.5.1 shows the commuting zones eligible for the relocation voucher. The targeted set consists of the 137 CZs in the bottom 30 percent of the 1880 manufacturing-employment share among CZs used in the quantitative analysis.

Figure A5.5.1: Commuting zones targeted by the relocation voucher



Notes: The figure shows the commuting zones used in the policy counterfactuals. Targeted CZs, shown in orange, are the 137 commuting zones in the bottom 30 percent of the 1880 manufacturing-employment share. Teal CZs are included in the quantitative analysis but are not targeted. Hatched CZs are not used in the quantitative analysis.

Counterfactual construction and outcome measures. To isolate the effects of the kin connections created by the relocation policy, I first solve the full policy equilibrium and impose its 1880 sector and location allocation in both economies. From 1910 onward, the comparison economy assigns each cohort the parental and extended-kin connections it would have inherited in the corresponding no-policy economy. Workers continue to value kin, while wages, prices, sector choices and location choices are re-solved endogenously. The two paths therefore have the same first-generation policy response and differ only in whether that response changes the kin connections inherited by later cohorts.

For outcomes reported for targeted families, I tag all workers living in $\mathcal{L}$ in 1880 and carry that tag forward to their descendants through the model's law of motion, so descendants remain in the targeted group wherever they subsequently live. Manufacturing and earnings effects are measured relative to the corresponding no-policy economy. Let N_t^q denote the number of CZ moves at model date t under policy path q , let N_t^0 denote the corresponding no-policy number and define the number of workers induced to move initially by $\Delta N_{1880} = N_{1880}^{\text{policy}} - N_{1880}^0$. Later migration through horizon T is reported per 100 initially induced movers as

$$\mathcal{R}_T^q = 100 \frac{\sum_{t \in \{1910, 1930, 1950, 1970\}, t \leq T} (N_t^q - N_t^0)}{\Delta N_{1880}}. \quad (\text{A5.5.5})$$

For the contribution of policy-induced changes in kin connections, I take the difference in $\mathcal{R}_T$ between the full policy path and this comparison path. Earnings use the CZ-sector wages generated by the corresponding equilibrium.

The welfare measure is earnings based. For each cohort I compute the change in log average earnings of targeted families relative to no policy. The later-generation measure in Figure 10 aggregates the 1930, 1950 and 1970 cohorts using the model's altruism parameter $\beta = 0.231$ and a stationary continuation after 1970:

$$W^{\text{later}} = \sum_{t \in \{1930, 1950, 1970\}} k_t \Delta \ln \bar{e}_t, \quad (k_{1930}, k_{1950}, k_{1970}) \propto \left(1, \beta, \frac{\beta^2}{1 - \beta}\right). \quad (\text{A5.5.6})$$

The scale of the weights is immaterial for the reported share of the policy effect. The 1880 tax is reported separately in Table A5.5.4 and is not part of this later-generation measure; the grant itself is excluded. Because the measure also omits amenities, migration costs and the direct utility value of kin proximity, it should not be interpreted as a complete welfare accounting of the relocation policy.

The value of kin proximity. The compensating-transfer exercise underlying Figure 9 uses the model's indirect utility to convert the difference in inherited horizontal-kin support between the origin and destination into dollars. For each candidate origin–destination corridor and sector k , I solve for G^{kin} such that

$$v(w_d^k + G^{\text{kin}}, p_d) - v(w_d^k, p_d) = \theta_H \psi_R^H [H^{\text{orig}}(o) - H^{\text{dest}}(o, d)]. \quad (\text{A5.5.7})$$

I allow G^{kin} to be negative when inherited kin support is greater at the destination than at the origin. Across candidate corridors, the median compensating transfer is \$157, the 75th percentile is \$249 and the 90th percentile is \$353. About 6 percent of corridors have a negative compensating transfer. The median is similar across high- and low-kin destination groups, at \$151 and \$162 respectively. The principal static cost captured by this exercise is therefore leaving origin kin rather than a mechanical difference in the destinations used in the targeting exercise.

Robustness across policy sizes. Table A5.5.2 reports the main outcomes for budgets between 0.25 and 1 percent of the 1880 wage bill. I also solve the relocation policy for budgets through 10 percent of the wage bill, financed by a proportional tax because a uniform lump-sum tax cannot finance budgets much above one percent; the qualitative pattern is unchanged. The table reports the values closest to the one-percent benchmark, over which the contribution of changes in kin connections is stable.

Policy-induced changes in kin connections account for 27–35 percent of the 1970 manufacturing

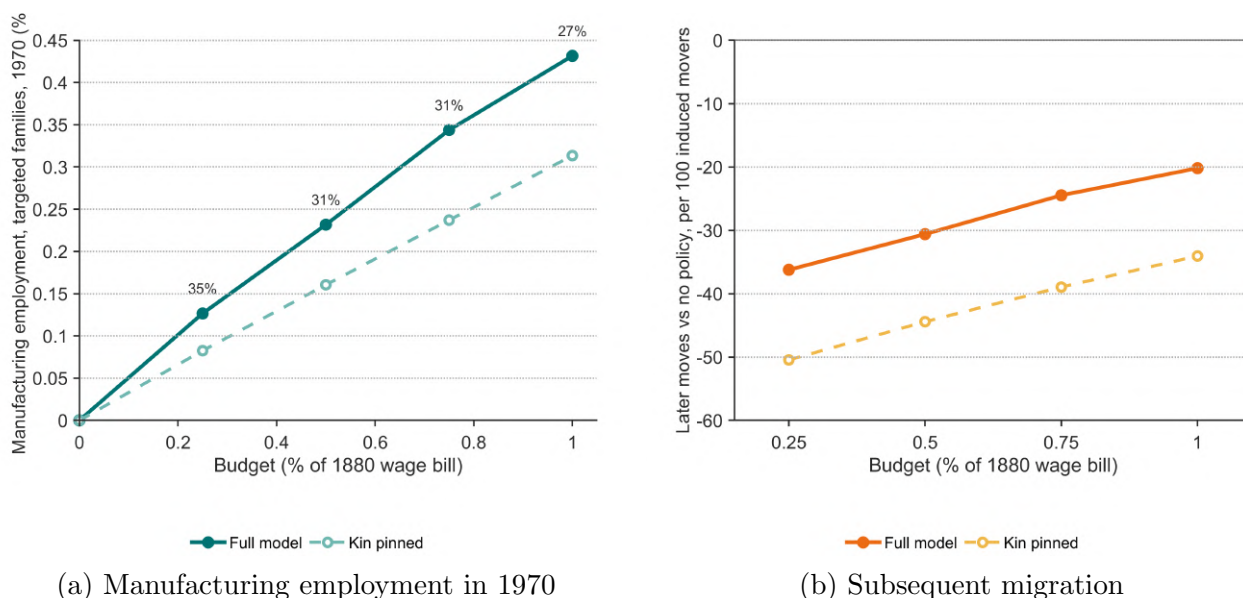
Table A5.5.2: Relocation-voucher effects across budgets

Budget (% of 1880 wage bill)	0.25	0.50	0.75	1.00
Induced movers in 1880 (000s)	4.7	9.2	13.7	18.1
<i>Manufacturing employment of targeted families, 1970 (%)</i>				
Full model	0.127	0.232	0.344	0.432
No policy-induced kin changes	0.083	0.161	0.237	0.314
Share from policy-induced kin connections (%)	34.8	30.7	31.0	27.3
<i>Later moves relative to no policy, per 100 induced movers</i>				
Full model	-36.2	-30.6	-24.5	-20.2
No policy-induced kin changes	-50.5	-44.4	-39.0	-34.1
Difference from policy-induced kin connections	14.2	13.8	14.5	13.9
<i>Earnings of targeted families, 1970 (%)</i>				
Full model	0.058	0.102	0.149	0.182
No policy-induced kin changes	0.045	0.087	0.127	0.167

Notes: “No policy-induced kin changes” imposes the full-model 1880 allocation but assigns subsequent cohorts the parental and extended-kin connections from the corresponding no-policy economy. Manufacturing and earnings are percent changes relative to the corresponding no-policy economy. Later moves are cumulative CZ moves from 1910 through 1970 relative to no policy, divided by the number of additional movers induced in 1880.

effect across these budgets and generate about 14 additional later moves per 100 first-generation induced movers.

Figure A5.5.2: Policy size and changes in kin connections



Notes: Panel (a) reports the 1970 manufacturing-employment effect for targeted families. Labels report the share of the full policy effect attributable to policy-induced changes in kin connections. Panel (b) reports cumulative moves from 1910 through 1970 relative to no policy per 100 movers induced in 1880. The difference between the two lines isolates the additional later migration generated by these changes.

Kin channels. I use a Shapley decomposition to separate the contribution of parental connections from extended kin and then divide the extended-kin component between location and sector connections. Let F denote the full-policy outcome, E the outcome when extended-kin connections from 1910 onward are replaced by their no-policy values, P the outcome when parental connections are replaced and B the outcome when both are replaced. The parental and extended-kin contributions are

$$\Delta^P = \frac{1}{2} [(F - P) + (E - B)], \quad (\text{A5.5.8})$$

$$\Delta^E = \frac{1}{2} [(F - E) + (P - B)]. \quad (\text{A5.5.9})$$

I split Δ^E between the location and sector components of extended kin using the analogous two-factor Shapley decomposition. Each arm preserves the same 1880 policy allocation and replaces only the indicated connections from 1910 onward, so the components measure how the kin environment created by the initial policy response changes later outcomes. Table A5.5.3 reports the decomposition across budgets.

Table A5.5.3: Kin channels of the voucher effect on manufacturing in 1970

Budget (% of 1880 wage bill)	0.25	0.50	0.75	1.00
Parents	0.014	0.014	0.025	0.006
Extended kin	0.033	0.068	0.086	0.110
Location	0.020	0.046	0.052	0.079
Sector	0.013	0.022	0.035	0.031
Total kin contribution	0.048	0.082	0.111	0.116

Notes: Contributions to the percent effect of the relocation voucher on manufacturing employment of targeted families in 1970. Parents and extended kin are obtained from a Shapley decomposition that replaces the corresponding connections with their no-policy values from 1910 onward. The extended-kin component is further decomposed into location and sector connections. Components may not add exactly because of rounding.

Extended kin account for most of the contribution from policy-induced changes in kin connections. At the benchmark budget, they contribute 0.110 points compared with 0.006 from parents, with about 72 percent of the extended-kin contribution operating through location connections.

Welfare across generations. Table A5.5.4 reports the earnings-based welfare effect separately by generation. Because the initial policy response is held constant, policy-induced changes in kin connections do not affect the initial tax burden and have little effect on the first post-policy cohort. Differences between the full and comparison paths emerge in later generations.

Table A5.5.4: Earnings-based welfare by generation

Budget (% of 1880 wage bill)	0.25	0.50	0.75	1.00
<i>Full model</i>				
1880: tax	-0.571	-1.146	-1.724	-2.305
1910	0.330	0.629	0.897	1.156
1930	0.218	0.433	0.649	0.849
1950	0.004	0.014	0.023	0.043
1970	0.058	0.102	0.149	0.182
<i>No policy-induced kin changes</i>				
1880: tax	-0.571	-1.146	-1.724	-2.305
1910	0.332	0.629	0.897	1.150
1930	0.197	0.399	0.595	0.789
1950	-0.004	-0.013	-0.016	-0.017
1970	0.045	0.087	0.127	0.167

Notes: Changes in log earnings of targeted families relative to no policy. The 1880 entry is the earnings effect of the lump-sum tax financing the relocation voucher. The grant itself is excluded. These are earnings-based measures and do not include the direct utility value of kin proximity.

At the benchmark budget, every post-policy cohort has higher earnings than under no policy. Differences between the full and comparison paths are small in 1910 and emerge more clearly in later generations.

Destination kin and targeting. The destination exercise varies where the relocation voucher can be used while holding the origin set and fiscal budget constant. The relevant unit is an origin–destination corridor (o, d) , where $o \in \mathcal{L}$ is the worker’s 1880 origin and $d \notin \mathcal{L}$. Objects that vary by sector or ancestral type within an origin are aggregated using the 1880 population weights of residents of that origin.

For each origin, let $P^0(o, d)$ denote the model’s 1880 probability of migrating to destination d when extended-kin connections are removed. This probability is determined by bilateral migration costs and destination wages, amenities and prices but not by observed extended-kin connections or policy outcomes. Let $H^{\text{dest}}(o, d)$ denote the inherited extended-kin component of the destination-specific amenity in equation (12), aggregated to the origin–destination level using the same 1880 origin weights. I define the candidate set $C(o)$ as the quarter of outside destinations with the highest $P^0(o, d)$. I then make the grant available at half of these candidate destinations. The high-kin rule selects the half with the most inherited destination kin and the low-kin rule the half with the least. Thus both rules are restricted to destinations that rank relatively highly on predicted migration before kin are used to separate them. They are not mechanically matched on $P^0(o, d)$ within the candidate set.

Table A5.5.5: Relocation vouchers at high- and low-kin destinations

	High kin	Low kin	Ratio
Mean inherited destination kin, H^{dest}	0.01195	0.00109	10.9
Grant, common budget (\$)	913	1,922	2.11
Induced movers (000s)	20.4	6.7	3.06
Fiscal cost per induced mover (\$)	5,852	17,929	3.06
Later moves per 100 induced movers, through 1950	10.1	2.9	3.48
Share of 1950 manufacturing effect from policy-induced kin connections (%)	45.9	5.5	–
<i>Equal number of first-generation movers</i>			
Grant (\$)	355	1,917	5.40
Budget (\$ million)	35.5	118.8	3.35
Induced movers (000s)	6.66	6.63	–

Notes: Both rules begin from destinations in the top quartile of predicted migration without extended kin. High- and low-kin destinations are the upper and lower halves of this set according to inherited destination kin. The groups are not mechanically matched on predicted migration within this candidate set. The common-budget comparison spends one percent of the 1880 national wage bill and recalibrates the grant separately under each rule. Induced movers are additional 1880 movers relative to the no-policy endogenous economy. Later moves report the additional post-1880 migration generated through policy-induced changes in kin connections, per 100 initial induced movers. In the equal-mover comparison, the grants are instead calibrated to produce approximately the same number of induced 1880 movers, so budgets differ.

Under a common budget, high-kin targeting induces about three times as many movers at roughly one third of the fiscal cost per induced mover. It also generates more than three times as many later moves per initial mover, while policy-induced kin connections account for 46 percent of the 1950 manufacturing effect compared with 6 percent under low-kin targeting. When the two rules instead induce the same number of first-generation movers, low-kin targeting requires 3.35 times the budget.