Networks and Geographic Mobility: Evidence from World War II Navy Ships*

Allison Green†

Click here for the most recent version

October 31, 2024

Abstract

This paper uses random assignment to Navy ships during World War II as a natural experiment to study how personal networks shape migration patterns. Using newly constructed data on 1.4 million sailors, I measure exposure to geographically diverse shipmates and estimate its impact on post-war migration. A one standard deviation increase in a sailor's exposure to shipmates from different states raises the probability of out-migration from his own state by 4%-5% by 1950. Effects on directed migration are larger but heterogeneous by destination, increasing moves to fast-growing Census divisions by over 15%. I then estimate a discrete choice migration model with embedded networks, revealing Navy ties encouraged long-distance moves, in part substituting short-distance moves that would have otherwise occurred. Using variation from Navy networks to construct instruments for the probability of migrating, I estimate large returns to network-facilitated migration, suggesting Navy ties enabled moves to higher-opportunity areas.

Keywords:

JEL codes: L1, N4, N9 O1, R2

*I am grateful to Kevin Liu, Logan Liguore, and Jennifer Green for excellent research assistance. I am also grateful to Eric Kilgore and the archivists at the National Archives Center in St. Louis for archive access and support. I thank my committee members Leah Boustan, Ilyana Kuziemko, and Stephen Redding for their invaluable support and guidance. Thanks also to Elena Aguilar, Kaan Cankat, Pier Paolo Creanza, Ellora Derenoncourt, John Fitz-Henley, Amy Kim, Mahsa Khoshnama, Casey McQuillan, Jessica Min, Rachel Moore, Sebastian Roelsgaard, Henry Shim, Carol Shou, Ruairidh South, and Owen Zidar for their helpful comments. Additionally, I am grateful to the participants of the Princeton Labor Workshop and the Yale Economic History Seminar. Financial support from the Princeton Industrial Relations section is also gratefully acknowledged.

†Princeton University: `aegreen@princeton.edu`

1 Introduction

A growing body of research has established a link between economic opportunity and where people live (Chetty et al., 2014). Yet, geographic mobility in the United States is low: 80% of adults under the age of 26 live within 100 miles of where they grew up (Hendren, Sprung-Keyser, and Porter, 2022). One potential explanation for limited geographic mobility, despite evidence of significant economic returns from migration, is the prohibitive cost of moving without an established network (Munshi, 2020; Blumenstock, Chi, and Tan, 2023). Most Americans lack out-of-state connections, with geographically broad networks typically concentrated among higher-income and more educated individuals (Chetty et al., 2022). Using one of the largest natural experiments in network formation in US history–Navy service during World War II, this paper assesses whether expanding the geographic scope of networks can increase migration and expand economic opportunity.

World War II provides an attractive empirical setting for exploring the role of networks in shaping migration patterns. The war exposed a substantial portion of young American men to individuals from outside their immediate communities, often for the first time: nearly 80% of all white men born between 1920 and 1926 served in the war in some capacity.[1] Given this scale, the networks formed through wartime service may have had profound effects on mid-20th-century America, fostering exposure to new ideas, diverse backgrounds, and geographically distant locations.

In this paper, I measure the impact of wartime networks on the migration decisions of rank-and-file sailors serving on Navy ships during World War II. This population was young, usually native-born, and without post-secondary education.[2] In part due to low migration during the Great Depression, the vast majority were living close to where they were born prior to entering naval service.[3] These characteristics make rank-and-file Navy sailors well-suited for studying the potential benefits of expanded geographic diversity in personal networks, as pre-existing research has shown migration gains are typically largest for young movers (Chetty, Hendren, and Katz, 2016; Nakamura, Sigurdsson, and Steinsson, 2022).

To estimate the impact of ship networks on migration, I leverage conditional random assignment to Navy ships. According to contemporaneous institutional records, assignments to ships were largely random. However, due to operational constraints in transporting personnel, men were more likely to serve on ships with others from similar geographic origins. The

[1]The military was racially segregated throughout World War II. In particular, the Navy was segregated by unit and occupation. This meant that except for a few small cases, Black men and other men of color in the Navy were not serving in combat roles on ships and are thus excluded from the scope of this paper.

[2]Men with college degrees and men enrolled in college were almost always assigned to officer positions.

[3]Figure A.1 shows that less than 15% of men born in the 1920s (largest World War II cohort) moved across state lines during childhood.

empirical strategy addresses this source of endogeneity by comparing migration outcomes for individuals from the same state who, due to the random chance of ship assignment, were exposed to different sets of peers during their service. This natural experiment, replicated across a large sample of sailors, enables analysis of how newly-formed connections shaped post-war geographic mobility.

I construct a novel dataset from archival records on the near-universe of ships active during the war. By processing 6.5 million scans of Navy personnel documents, I create a dataset that includes detailed information on over 1.4 million sailors across more than 5,000 ships. For each sailor, I document their precise dates of service and pre-war state of residence. This information allows me to identify not only which sailors served together on specific ships at the same time, but also the geographic diversity of shipmates that each individual was exposed to during service.

Using this dataset and empirical framework, I estimate the impact of these wartime networks on two key dimensions of migration: out-migration and directed migration. Out-migration measures movement away from one's own pre-war residence, capturing overall geographic mobility. Directed migration examines movement toward the pre-war residences of fellow shipmates, reflecting the influence of specific network ties. Together, these two margins reveal both the extent to which direct ties formed during service influenced specific migration destinations and the aggregate network effect of shipmates on general mobility.

I find that shipmates significantly influence both whether a Navy sailor moves (out-migration) and where they move (directed migration). By 1950, a one standard deviation increase in exposure to out-of-state shipmates raises the likelihood of moving out-of-state by 4.2%. The effect is larger for long-distance moves: a one standard deviation increase in exposure to out-of-region shipmates increases the likelihood of moving out-of-region by 6.5%. The impact of shipmates on directed migration is even more pronounced, though heterogeneous across destinations. A one standard deviation increase in exposure to shipmates from fast-growing states in the Pacific increased migration to those areas by over 15% by 1950. Notably, Navy networks had little impact on migration to slower-growing regions like the Midwest, suggesting these networks primarily facilitated moves to areas of high economic opportunity.

To fully leverage variation in ship network composition, I extend the analysis by developing a discrete choice migration model that incorporates network ties formed through Navy service. By embedding network effects into a discrete choice model, I evaluate Navy ties to a particular state within the broader context of all available network connections on a ship. The model includes heterogeneous network effects between states, capturing various channels through which networks influence migration. Notably, the model allows the value of additional network ties to vary based on the pre-existing density of networks in each destination. Using random variation from Navy assignments to estimate the model, I compute counterfactuals that decompose the role of Navy networks in explaining observed migration patterns.

The counterfactual exercises quantify the heterogeneous effects of networks formed during World War II on post-war interstate migration patterns. On average, a 10 percentage point increase in exposure to shipmates from California, the fastest growing state in the 1940s, increases migration to California by 23%. However, effects by origin state range from 7%-37%, with the largest increase observed for people from the East coast who are unlikely to have pre-existing ties to Western states. Examining total migration effects, individuals from the Midwest were proportionally most responsive to these wartime networks, with their moves primarily directed toward high-growth states like Florida, Texas, and California. A back-of-the-envelope calculation suggests that over 14% of migration to California between 1945 and 1950 can be attributed to exposure through wartime networks. Collectively, these counterfactual exercises indicate that networks formed during military service were a major force in driving migration to high-growth areas.

I then explore how network formation on ships, particularly through shared ethnicity between shipmates, influences post-war migration. Shared characteristics between sailors may strengthen social bonds and therefore amplify the impact of shipmate connections on migration decisions. By 1950, a higher share of co-ethnic shipmates increases both the overall likelihood of migrating and the probability of moving to specific states where these co-ethnics are from, with increased exposure to co-ethnic shipmates being about 2.5 times more influential than shipmates of a different ethnicity. However, the influence of co-ethnic largely disappears in the long-run, suggesting that the salience of ethnic connections diminishes over time.

The final section of the paper examines whether moves facilitated by Navy networks led to greater economic opportunity. Using random variation in networks and the estimated discrete choice model, I construct instruments for the probability a Navy sailor moves out-of-state, out-of-region, and to a state in the Pacific Census division. I then estimate the returns to earnings for men induced to move due to their wartime networks using a two-stage least squares estimation procedure. These estimates can be interpreted as the local average treatment effect for compliers–those whose migration decision was marginal to the influence of Navy networks. I find economically large impacts of network-induced moves on lifetime earnings: individuals induced to move out-of-state by their Navy networks reside in zip codes with 59% higher income by the time of their death. While substantial, these estimates are credible within the context of the literature on returns to migration. The findings suggest that networked migration facilitated by Navy service allowed this population of men to move to areas of greater economic opportunity.

This paper makes several contributions to existing literature. First, this work advances our understanding of networks' role in shaping migration decisions. Theoretical models predict that networks increase migration rates by lowering migration costs (Carrington, Detragiache, and Vishwanath, 1996) and providing destination-specific information (Porcher, Morales, and Fujiwara, 2024; Porcher, 2020; Granovetter, 1973). Empirical evidence confirms that origin-destination links increase international migration rates (Beine, Docquier, and Özden, 2011; Dolfin and Genicot, 2010; McKenzie and Rapoport, 2010; Munshi, 2003). In the context of internal migration, Blumenstock, Chi, and Tan (2023) utilize phone data in Rwanda to show that individuals are more likely to migrate to areas where they have stronger network ties. Büchel et al. (2020) find that networks account for a substantial portion of idiosyncratic location choices in Switzerland. Kinnan, Wang, and Wang (2018) use a historical policy that generated random links between rural and urban Chinese provinces, revealing persistent migration patterns. In American settings, Stuart and Taylor (2021) employ birth town as a proxy for migration networks in 20th-century migration events, and Costa et al. (2018) demonstrate that Civil War veterans from the same companies are more likely to live in close proximity decades after the war. In recent work, Koenen and Johnston (2024) measure networks using Facebook data and employ exogenous differences in the timing of moves to show that recent college graduates are more likely to move to Commuting Zones where network ties are located.[4]

Although these studies establish a strong association between networks and migration, identifying a precise causal link remains challenging due to the scarcity of quasi-experimental variation in individual networks. Most existing empirical research use origin home communities as proxies for networks and then exploit random variation in the spatial dissemination of people from home networks to measure their role in facilitating migration.[5] However, plausibly exogenous variation in individual networks is rare, particularly regarding shocks to the size and composition of personal networks.[6] In my paper, I leverage random extensions to individual networks through naval service in World War II. By examining how these wartime connections influenced post-war migration within the United States, this study provides a

[4] I note that this paper and my own were developed separately and without knowledge of each other.

[5] Two landmark papers in this style are Munshi (2003), which uses rainfall as an instrument to measure random variation in the number of individuals from a Mexican origin community located in different U.S. communities, and Beaman (2012), which uses variation in refugee resettlement locations to proxy for access to pre-existing networks.

[6] Both Blumenstock, Chi, and Tan (2023) and Koenen and Johnston (2024) measure personal networks using phone records and Facebook data respectively. They each take the network as fixed and employ empirical strategies to measure random variation in where individuals from their network live.

unique opportunity to isolate the causal impact of network formation on migration decisions.

Second, this paper contributes to the literature on returns to migration by introducing a novel approach to estimating these returns. Previous studies have primarily relied on three methods. The first uses displacement events as natural experiments, with several studies finding positive earnings effects from forced relocations - whether from natural disasters (Nakamura, Sigurdsson, and Steinsson, 2022; Deryugina, Kawano, and Levitt, 2018), wartime displacement (Sarvimäki, Uusitalo, and Jäntti, 2022), or government internment (Arellano-Bover, 2022). The second approach controls for family background by comparing migrant and non-migrant siblings. This method has been particularly influential in studying the Great Migration, where researchers identified substantial returns for both Black and white men who left the South (Collins and Wanamaker, 2014; Boustan, 2016). Related work using sibling comparisons has found moderate positive effects of internal migration on early career earnings more broadly (Ward, 2022). This paper extends this literature by employing random variation in network exposure as an instrument for migration, offering a novel method to estimate the causal effects of migration on economic outcomes. Consistent with studies that find the largest returns to migration for younger movers (Nakamura, Sigurdsson, and Steinsson, 2022; Chetty, Hendren, and Katz, 2016), the analysis focuses on rank-and-file Navy personnel who were typically in their early twenties during service.

The last approach to studying returns to migration uses evidence from policies designed to subsidize migration. The most prominent example is the Moving to Opportunity (MTO) experiment, which subsidized moves from high- to low-poverty neighborhoods (Chetty, Hendren, and Katz, 2016; Ludwig et al., 2013; Clampet-Lundquist and Massey, 2008; Kling, Liebman, and Katz, 2007). While Chetty, Hendren, and Katz (2016) find positive effects for children who moved young, the program showed no earnings gains for the average beneficiary. Barnhardt, Field, and Pande (2017) find housing lottery winners in India experienced little economic gains and greater social isolation after moving. These moving subsidy programs consistently face low take-up rates, which Bergman et al. (2024) attribute partly to missing destination networks. These papers suggest network barriers may limit the effectiveness of subsidized migration programs by reducing both take-up rates and participants' ability to access economic opportunities in new locations. The analysis of Navy networks facilitating moves to high-opportunity areas provides new evidence on mechanisms that shape the impact of geographic mobility programs.

Third, my paper extends the literature on peer effects and networks, as summarized by Bramoullé, Djebbari, and Fortin (2020), with a particular focus on network formation in early adulthood through randomized peer interactions. Research in college and military settings is especially relevant. In college contexts, Sacerdote (2001) examines the effects of randomly assigned dorm roommates, and Michelman, Price, and Zimmerman (2022) analyzes the impact of dorm assignments at Harvard in the 1920s, and Shue (2013) studies networks formed among Harvard MBA students. In military environments, Einiö (2019) demonstrates that Finnish soldiers in military dorms earn higher incomes when exposed to higher-income peers. Conversely, Carrell, Sacerdote, and West (2013) find negative outcomes from an optimal peer mixing experiment at the Air Force Academy, attributed to endogenous peer formation. Costa and Kahn (2003) observe higher desertion rates in more heterogeneous Civil War battalions, and Guo, Jackson, and Jia (2024) show that increased exposure to West Point cadets from Northern states makes a person more likely to join the Union over the Confederacy. These studies generally yield mixed results on the benefits of peer mixing, with effects often attenuated when cross-group differences are prominent. This paper advances this literature by examining the impact of randomized peer exposure during World War II and investigating how factors such as class, ethnicity, and geographic origin influence network formation and mediate treatment effects in this unique historical context. This paper is also one of few papers that focuses on the role of randomized peer exposure on geographic mobility.

Finally, this paper contributes to the literature on World War II's impact on the U.S. post-war economy. This research primarily encompasses two main strands: (1) the effects of war mobilization and demobilization on aggregate economic activity (Fishback and Cullen, 2013; Fishback and Jaworski, 2016; Moretti, Steinwender, and Van Reenen, 2023; Garin and Rothbaum, 2024), and (2) the war's influence on human capital accumulation and wages (Bound and Turner, 2002; Collins and Zimran, 2024; Althoff and Szerman, 2024; Bedard and Deschênes, 2006; Acemoglu, Autor, and Lyle, 2004; Aizer et al., 2020). This study distinguishes itself by being among the first to utilize ship-level variation in studying World War II, building on Suandi (2022) use of submarine promotion data. By leveraging granular ship-level data to measure the effect of wartime networks on subsequent economic outcomes, I provide novel insights into the long-term consequences of wartime social connections.

The paper proceeds as follows: Section 2 describes aggregate trends in geographic mobility and military service in relation to World War II. Section 3 describes the construction of the core World War II Navy dataset. Section 4 describes the empirical strategy and presents results on the causal impact of Navy networks on migration. Section 5 presents the discrete choice model and conducts counterfactual exercises to quantify the impact of wartime connections on aggregate migration patterns. Section 6 explains the role of ethnic ties on network formation aboard Navy ships. Section 7 discusses the construction of a migration instrument based on random variation in Navy networks, and uses the instrument to estimate the returns to network-facilitated migration. Section 8 concludes.

2 Historical Background

2.1 Relationship between migration and military service

Geographic mobility for young, white American men peaked in the mid-20th century, coinciding with high rates of military participation during World War II.[7]

Figure 1 illustrates this trend, showing the share of men in their 30s living outside their state of birth, controlling for childhood moves.[8] The figure reveals a sharp rise in young-adult migration (ages 20-40) in the first half of the 20th century, peaking for those born in the 1920s—the cohort with the highest level of World War II service. Cross-state prime-age geographic mobility has slowly declined since. Then, overlaying military participation rates for 20th-century birth cohorts shows a striking correspondence between geographic mobility and military service rates.

Military participation remained high between World War II and Vietnam due to the draft (1940-1973). Since the end of conscription, the share of white men joining the military has decreased steadily, with modern enlistment rates below 10% for young white men. The tight relationship between military participation and aggregate geographic mobility is supported by research showing that veterans are more geographically mobile than non-veterans, even when controlling for selection into service (Bailey, 2011).

This peak in mobility for the World War II cohort aligns with broader trends in U.S. internal migration throughout the 20th century. Rosenbloom and Sundstrom (2004) document that interstate migration rates reached their highest levels in the mid-20th century, up from their lowest point around the turn of the 20th century. The elevated level of mobility persisted for several decades after the war but has been declining over the past 50 years (Molloy, Smith, and Wozniak, 2011). Research by Hendren, Sprung-Keyser, and Porter (2022) on contemporary migration trends indicates that 65% of young Americans now live in the same commuting zone where they grew up, and 80% live within 100 miles of their childhood home.

World War II catalyzed large-scale migration flows throughout the United States, extending beyond those directly involved in military service (Harriman, 1948). War production centers attracted significant numbers of workers deemed ineligible for military service, including women and men with occupational or physical exemptions (Johnson, 1994). The South, particularly agricultural areas, experienced substantial out-migration as individuals

[7] See Rosenbloom and Sundstrom (2004) for a discussion of aggregate cross-state migration trends. Historical trends in long distance geographic mobility differ significantly for Black men compared to white men (Hall, 2009).

[8] The figure shows the change in probability of living outside state of birth from ages 10-19 to 30-39.

relocated to western states with high concentrations of defense production. By spring 1943, the influx of migrant workers to California's defense industry was so substantial that the San Francisco Chronicle dubbed it "The Second Gold Rush" (Johnson, 1994). However, many of these wartime migrants ultimately returned to their origin communities after the war ended (Harriman, 1948).

Contemporaneous reports indicate that World War II veterans' were highly geographically mobile after the war. According to a 1947 Census report, only 30% of World War II veterans maintained their 1940 residential address by 1947 (Harriman, 1948). Former farm residents showed the highest propensity to relocate between 1940 and 1947 though most of these moves occurred within their original county (Harriman, 1948).

Several mechanisms may explain the link between military service and increased geographic mobility. First, World War II veterans were entitled to the benefits of the Service Readjustment Act of 1944, more commonly known as the GI Bill. The GI Bill provided veterans access to low-rate mortgages and protection against default, helping accelerate homeownership rates throughout the mid-20th century (Fetter, 2013). These benefits coincided with a severe housing shortage in the late 1940s, driven largely by low housing construction during the Depression and war years (Hausser and Jaffe, 1947). To address these shortages, particularly for World War II veterans, there were large public and private investments in new housing construction, much of it centered in newly developing suburbs.

Anecdotal evidence suggests social connections formed during military service also played a role in post-war migration patterns. For instance, Major Richard Winters, a member of the Band of Brothers, notes in his memoir how he received a job through one of his commanding officers: "Within two weeks of returning home [from World War II], I accepted Lewis Nixon's invitation to travel to New York City and meet his parents. His father offered me a job and in January 1946, I became personnel manager for the Nixon Nitration Works in Nixon, New Jersey."

Similarly, Walter W. Schumacher, who served aboard the light cruiser USS Omaha, describes in the Veteran History Project how a friendship formed during basic training influenced his decision to move from New York to Toledo, Ohio: "I met [my wife's] brother in the North Atlantic Great Lakes Service School, so I come to Toledo a few times with him [after the war]. And so [me and my wife] got together...in 1946 we got married in New York. She didn't want to leave Toledo, so it didn't make no difference to me [so we moved to Toledo]." These examples demonstrate two distinct mechanisms through which military service shaped migration: network-based job referrals that provided economic opportunities in new locations and social ties that generated non-pecuniary reasons to move through marriage and family connections

2.2 US Navy and World War II

The US Navy underwent a dramatic transformation during World War II, evolving from a peacetime force of around 100,000 volunteers in the 1930s to a massive wartime fleet of over 3.5 million personnel.[9] This rapid expansion was set in motion even before the United States officially entered the war. In July 1940, responding to growing global instability, Congress passed the Two-Ocean Navy Act, authorizing a significant increase in naval personnel and funding an extensive shipbuilding program (Morison, 1963).

The buildup of naval forces was closely tied to the implementation of conscription. The first peacetime draft registration in September 1940 required men between the ages of 21 and 45 to register for potential military service. By the time of the Pearl Harbor attack on December 7, 1941, which precipitated U.S. entry into the war, the Navy had already grown to 325,000 personnel. Over the next four years, this number would increase more than tenfold.

The composition of Navy recruits changed over the course of the war. Initially, the Navy's ranks grew through a combination of the draft and voluntary enlistment. However, in December 1942, voluntary enlistment was suspended.[10] This decision was made to maintain a sufficient agricultural workforce and to alleviate tensions between the Army and Navy, as the Navy's stronger recruiting apparatus had been attracting a disproportionate share of volunteers.

By 1945, the U.S. Navy had reached its largest size in American history with 3.4 million active personnel, of whom approximately 1.5 million were enlisted men serving on ships.[11] who served aboard Navy ships.[12] While Navy personnel were deployed across various theaters, the Pacific was the primary focus, with over 80% of overseas personnel stationed there by August 1945.[13] Despite the rapid expansion and the dangers of wartime service, casualty

[9] Source:https://www.history.navy.mil/research/library/online-reading-room/title-list-alphabetically/h/history-of-the-us-navy/personnel-strength-1794-1990.html. Last Accessed: 10/11/24

[10] Ended via Executive Order 9279, which closed voluntary enlistment for men aged 18-37. The Navy still maintained an active recruiting arm for enlisting 17-year-olds into voluntary service.

[11] Within Navy terminology, they use the term "enlisted" to refer to all non-officer personnel regardless of whether they volunteer or were drafted.

[12] Other jobs outside of active duty ship service include, domestic shipbuilding and administrative work, overseas shore service, aerial service, etc.

[13] As of August 31, 1945, the Pacific theater (both ashore and afloat) accounted for approximately 82.5% of peak Navy personnel strength across major theaters, with 1,366,716 personnel. The North Atlantic theater had 9.1% (150,046 personnel as of June 30, 1944), the Mediterranean 5.4% (90,175 on August 31, 1944), and continental Europe 3.0% (49,801 on November 30, 1944).Source: https://www.history.navy.mil/research/library/online-reading-room/title-list-alphabetically/u/us-navy-personnel-in-world-war-ii-service-and-casualty-statistics.html. Last Accessed: 10/11/24

rates among enlisted personnel were relatively low: out of over 3.5 million who served, 32,925 were killed and 34,478 wounded.

The Navy's fleet expansion mirrored its personnel growth, growing from 478 vessels in 1940 to 6,768 by August 15, 1945 (Victory in Japan day). Navy ships varied significantly in size: 56% were small (under 100 people), 37% medium (150-400 people), and 7% large (over 1000 people). Despite size differences, ships shared a common organizational structure characterized by a vertical hierarchy of officers and enlisted personnel, with rank-and-file making up 70-80% of the crew. Horizontally, ships were organized into functional units like deck, engine room, and mess.

The end of the war brought about a massive demobilization of Navy personnel. At the end of the war, most of the rank-and-file personnel were released back into civilian status. Men were returned to a major port or Navy station, and given transportation back to their pre-war residence. Using a sample of around 10,000 individual separation documents, I find that over 90% of all enlisted sailors report intending to return to the same address they were living at prior to the war upon immediately exiting the Navy.[14]

Despite the large-scale demobilization, many veterans maintained ties with the Navy. A substantial portion of men who served in the Navy, even those who were initially drafted, chose to enter the Navy reserves and approximately 25% of men who served in the Korean War also served during World War II.[15]

3 Data

To study the long-run impact of network ties formed in the U.S. Navy during World War II, I assemble a dataset that combines detailed information on individuals' service histories with measures of their economic and geographic outcomes after the war.

The primary data source is a newly constructed collection of World War II Navy Muster Rolls, which contain quarterly snapshots of the full roster of enlisted personnel for the universe of Navy ships active during the war. Using these records, I identify the specific ships and time periods of each sailor's service, enabling the construction of comprehensive individual-level shipmate networks. I then link these service records to data from the 1940 and 1950 U.S. Censuses and death records to track geographic mobility patterns over a person's lifetime.

The resulting dataset contains complete World War II Naval service histories and pre-

[14]Each sailor filled out a separation document (NAVPERS-563) upon exiting the Navy. This document contained address prior to the war, and the address for which the person registered with the selective service board, and the address the person intends to seek employment. I collected over 10,000 separation documents at National Archives Site in St. Louis

[15]Source: https://www.va.gov/vetdata/docs/specialreports/kw2000.pdf

and post-war outcomes for approximately 300,000 individuals. This section details the data construction process, including the digitization and cleaning of the Muster Rolls, the record linkage methodology, key variables used in the analysis, and summary statistics of the final sample.

3.1 World War II Navy Muster Rolls

The primary data source is a collection of World War II Navy muster rolls, reports submitted quarterly by each U.S. Navy ship to the Bureau of Navy Personnel between 1939 and 1949.[16] For each muster roll, I focus on information from two document types: (1) quarterly censuses of every enlisted sailor on the ship (5-10% of scans), and (2) monthly reports of any personnel changes, such as sailors boarding or leaving the ship, or changes in their rating (20-25% of scans).[17]

I digitize 6.5 million scanned images of Muster Rolls from the National Archives Catalog using optical character recognition (OCR) and LayoutParser.[18] Figure A.3 Panel A presents examples of the quarterly census and monthly personnel change forms.[19] Since the original documents vary in quality, the digitized data contains encoding errors and missing fields. To address these issues, I use sailors appearing across multiple scans over their service period. By combining information across all records pertaining to a given individual, I am able construct accurate service histories for a much larger sample of individuals than if I relied on individual entries alone. Figure A.3 Panel B illustrates this cleaning procedure by showing all records pertaining to a particular sailor "Hugh Berry", including OCR errors. The final row displays the corrected service record after combining information across observations. On average, each sailor's service number–a unique seven-digit identifier assigned by the Navy–appears on 8.5 distinct scans, providing ample opportunity to correct errors. Further details on the data cleaning and construction are discussed in Appendix B.1.

Figure A.3 Panel C provides summary statistics for the cleaned sample. I identify over 1.4 million unique rank-and-file sailors who served across over 5000 ships between 1941 and 1945. By historical account, around 1.5 million rank-and-file personnel served on Navy ships

[16] All formal "activities" were required to submit these quarterly reports. An activity is defined as "a unit, organization, or installation performing a specific mission or function and established under a commanding officer, officer in charge, etc.; e.g., naval air station, naval shipyard, naval station, a specific air squadron, ship, etc."

[17] The most common personnel changes are individuals boarding the ship ("Received"), exiting the ship ("Transferred"), and changes of rating ("Change of Rating"). Less common types of personnel changes include individuals going AWOL, hospitalizations, and short leaves of absence.

[18] It is important to note that not all 6.5 million images contain usable data; some are cover pages, blank forms, or otherwise irrelevant to my analysis and are thus disregarded.

[19] Each quarterly census scan reports up to 35 sailors, while each monthly report of changes includes up to 15 individuals.

during the war, so I have fairly complete coverage of the universe of the population of interest. The modal sailor served on only one ship over the course of his service, and over 95% of sailors served on two or fewer vessels. The average person is in my sample for a period of 19 months, with the average stint on a given ship lasting 15 months.

3.1.1 Supplementary Military Data

I supplement muster roll data with additional military records to explore how varying peer interactions and service circumstances might impact outcomes. The *Dictionary of American Naval Fighting Ships* (DANFS), published by the Navy from 1959 to 1991 and later digitized by volunteers, offers detailed historical accounts for each vessel. These accounts include ship specifications, operational timelines, major engagements, casualty reports, and commendations. The basic details on operational dates and ship dimensions serve to validate the cleaned muster roll data. Moreover, the battle and conflict records enable analysis of cross-ship variation in combat intensity, potentially shedding light on how these experiences influenced the strength of bonds formed among crew members. I also use a wartime report from the Naval Health Research Center to measure the share of wounded, missing in action, and killed for each of the 150 distinct categories of ships.

To gain additional insights into the types of ties that a person might form within a ship, I use historical Navy Occupation Ratings data. This data provides information on each rating code (e.g., "EM3C"), full-title occupation (e.g., "Electrician's Mate 3rd Class"), pay grade (e.g., "III"), and rating branch (e.g., "Artificer's Branch"). By using this information, I can explore how the nature of relationships might vary based on occupational proximity and hierarchy. For instance, two people of the same rating are more likely to work together and thus may be more likely to form close ties. Similarly, the relative positions of individuals within the rating hierarchy could have different impacts on the types of ties formed.

Finally, I use archival internal documentation from the Bureau of Navy Personnel on how service numbers were allocated across enlistment centers to identify the place each person enlisted by their service number.

3.2 Measuring Characteristics of Shipmates

To analyze the relationship between shipmate characteristics and long-run migration outcomes, I construct networks of shipmates for each individual in the data.

3.2.1 Individual Characteristics

I first construct individual characteristics using data available in the Muster Rolls. These characteristics serve as proxies for geographic, economic, and demographic attributes, which I then use to construct network measures of each sailor's exposure on their ship. By utilizing information only available in the Muster Rolls, I can create network measures for the entire ship network, and not just the subset of individuals I subsequently link to other datasets.

I use place of enlistment as a proxy for pre-war geographic residence. During World War II, over 100 receiving stations were assigned specific service number ranges. For example, sailors enlisting in Buffalo, New York, received numbers between 8052000 and 8066999. I use these ranges to determine enlistment locations for the entire sample.[20]

To impute ethnicity/ancestry, I employ the method developed by Abramitzky, Boustan, and Eriksson (2020) using ethnic differentiation among names.[21] I compute indices from white men in the 1940 census, and construct these indices for the sixteen largest ethnicities/nationalities. I supplement this with an index for Jewish names index used in Abramitzky, Boustan, and Connor (2024). To impute pre-war education, income, and occupation, I use complete names and state of enlistment.[22] For all three economic characteristics, I impute the value belonging to the head of household of an individual with that name. I do so because a substantial portion of Navy sailors during World War II were still living at home in 1940, and it avoids confounding life-cycle effects with differences in economic status.

Panel C of Figure A.3 summarizes these imputation measures. I successfully identify the state of enlistment for over 97% of the sample, ethnicity for 96%, and occupational score for 89%. Appendix B.1 discusses each imputation method in more detailing and provides additional validation.

[20]Appendix Figure A.2 maps the enlistment locations, demonstrating their wide geographic distribution and supporting the use of place of enlistment as a proxy for pre-war residence. At least one enlistment center was active in each of the 48 contiguous states.

[21]This method calculates the likelihood of a name belonging to one of eight ethnic groups, assigning ethnicity when the score exceeds 0.7. The formula for Italian ethnicity, for example, is:

$$F_{nh} = \frac{\frac{\text{\# of People with last name n and Italian}}{\text{\# of people Italian}}}{\frac{\text{\# of People with last name n and Italian Father}}{\text{\# of people with Italian}} + \frac{\text{\# of People with name n not Italian}}{\text{\# of people not Italian}}}$$

I construct name indices by full name, first name, and last name.

[22]For instance, a John Smith from Ohio in my sample is assigned the average 1940 occupational score of all John Smiths in Ohio. For men whose complete full name and state do not match to the 1940 Census, I then impute characteristics based on surname and state of enlistment.

3.2.2 Computing Network Measures

Using these imputed characteristics, I then construct individual-level measures of their exposure network to other sailors on Navy ships. The primary focus is on the geographic network of exposure, which captures the extent to which a sailor was exposed to shipmates from different states. I supplement this with additional characteristics of shipmates including their occupational role on the ship and imputed ethnicity and income.

To measure geographic exposure to shipmates, I create a vector of 49 elements for each sailor. The first 48 elements correspond to exposure to people from the 48 contiguous states, while the 49th element represents an "other" category that includes territories and non-American enlistments.[23] Each element in this vector represents the duration-weighted share of shipmates from that particular state or category.[24] For sailors who served on multiple ships, I focus solely on their first ship assignment to construct these network measures. This approach reflects that while assignment to the first ship is conditionally random, assignment to subsequent ships is less so.[25]

To complement the geographic exposure measure, I incorporate data on shipmates' boarding times and occupational roles within the ship. The timing of when shipmates boarded relative to each other allows for exploration of how interaction duration might influence network formation and group dynamics. Using Navy rating codes, I identify each sailor's ship occupation, enabling the examination of hierarchical dynamics within the ship. Sailors with similar occupations likely had greater exposure to one another, potentially forming stronger network ties. This occupational data also facilitates analysis of how interactions between sailors in different hierarchical positions might vary. Finally, I use imputed characteristics on ethnicity, income, and occupation based on sailors' names and enlistment locations. By incorporating these imputed characteristics into the network measures, I can analyze how exposure to diverse backgrounds and socioeconomic statuses within the ship's network might influence network formation and post-war migration decisions.

3.3 Record Linkage

Linking individuals from Navy service records to three key periods in their lives – pre-war (1940 Census), prime-age (1950 Census), and death (Numident Social Security Records, Veteran Affairs, and FindAGrave) – forms a crucial part of this analysis. The Muster Rolls

[23] Because Alaska and Hawaii were territories until 1959, enlistments in both of these places are grouped into the "'other" category.

[24] Weighted and unweighted measures of ship characteristics have a correlation of 0.83, reflecting that the set of shipmates a person was serving with was relatively stable across time.

[25] Assignment to ships is discussed in full in Section 4. Future robustness checks will relax this assumption and pool exposure across ships.

provide limited characteristics for linkage, necessitating a multi-step record linkage procedure that maximizes link rates while minimizing false positive matches.[26]

Step 1: Restrict Data. I restrict the sample to men born between 1905 and 1928, covering World War II draft eligibility and active combat service requirements. Men known to have served in the Army are excluded.[27]

Step 2: Bilateral Links. I construct bilateral links between datasets using both deterministic and probabilistic approaches. The deterministic method, following Abramitzky et al. (2021), utilizes all common variables with non-missing values. For instance, when linking the Muster Rolls to the 1940 Census, I use first and last names, allowing for both exact and fuzzy matching, and verify uniqueness across datasets.[28]

Complementing this, I employ probabilistic record linkage using the Python package `splink`, which implements Fellegi and Sunter (1969). This approach is well-suited for cases with missing observations, measurement error, and continuous variables such as distance (Enamorado, 2021). For example, when linking the Muster Rolls to the 1940 Census, I use first and last names, birth year (upper bound in the Muster Rolls), middle initial, and the geographic distance between 1940 residence and enlistment city. To prevent false positives, I require all non-missing variables to match within a specified bandwidth and verify uniqueness across both datasets.[29]

Step 3: Harmonize Deterministic and Probabilistic Links. I harmonize links established through deterministic and probabilistic methods. Discrepancies between the two procedures, occurring in less than 0.5% of cases, are dropped from the analysis.

Step 4: Establish a Chain of Links. In this final step, I use the network of bilateral links to connect individuals across multiple data sources. This approach enables indirect links between datasets that lack a direct match, maximizing the use of available information. For example, I may link a sailor from the Muster Rolls to the 1940 Census via the Social Security Death records, using name and date of enlistment for the first link, and name and place of birth for the second.

[26] The Muster Rolls provide name, date of enlistment, and place of enlistment. Date of enlistment provides an upper bound on year of birth and facilitates linking to social security records. Place of enlistment aids in linking to 1940 Census residence.

[27] Over 99% of Navy enlistees during this period were born within this range, according to the Veterans Affairs index. Branch switching after formal enlistment was exceedingly rare.

[28] To account for common OCR encoding errors and the limited set of linking variables, names must be unique over a fuzziness bandwidth. For example, a name that is unique on exact match must also be unique within a Jaro-Winkler distance of 0.05.

[29] For instance, when linking between the 1940 Census and Social Security records, I use the first names of the person's mother and father when available. To avoid false positives between household members, I restrict links to those where the Jaro-Winkler distance between first names is less than 0.3 and birth years fall within a 3-year bandwidth.

The final linked sample comprises 478,000 sailors observed in both the muster rolls and the 1940 Census. Using the IPUMS 1940-1950 MLP Crosswalk, 266,000 of these individuals are linked to the 1950 Census. Additionally, 578,000 individuals are linked to at least one death location measure (either Social Security records or FindAGrave). The achieved match rates of 35%, 20%, and 42% respectively are at the higher end of match rates typically found in the literature. Appendix B.3 provides detailed information on each step of the linking process. Match rates for bilateral linkages and overall linkage across all datasets are reported in Appendix Table A.2.

3.4 Additional Data Sources

Full Count Censuses

The 1940 Census provides pre-war geographic, demographic, and household characteristics of men who served in the Navy during World War II. With a median age of 16 in 1940 for men in the linked sample of Navy men, the measured household characteristics represent a mix of parental household attributes (for those still living at home) and individual characteristics (for those who were household heads). I measure prime-age outcomes using the 1950 Full Count Census, which contains comprehensive data on residence, occupation, and household characteristics.[30] For measuring income in both Censuses, I use occupational score which reports the average income (in hundreds of 1950 dollars) for each 1950 occupational code.[31]

Location and Characteristics at Death

To measure individuals' outcomes over their lifetimes, I use mortality records from the Social Security Administration's Numident file as the primary data source. These records cover a high proportion of individuals who died between 1985 and 2007, totaling approximately 50 million records. The Numident file provides the zip code of residence at the time of death, which serves as the main measure for end-of-life location and economic outcomes.

To supplement this data and extend the sample outside of deaths covered by Numident, I collect data from the website "FindAGrave.com", the world's largest online database of gravestones. This publicly searchable database contains burial locations, demographic information including places of birth and death, and military service details including branch

[30] The 1950 Census asked fewer questions of all individuals than the 1940 Census. Questions on education and income were only asked for the sample-line which covered 20% of the population.

[31] Household income is reported only on the sample line in 1950, and there are current irregularities in the field in the public release of the 1950 Full Count Census. I use occupational score in both periods for consistency.

and conflict. After restricting to veterans born between 1900 and 1928 and excluding those identified as Army servicemen, the database yields 1.8 million individuals eligible for linking.

Finally, I use data acquired from a Freedom of Information Act (FOIA) request to the Veterans Affairs Bureau.[32] These data contain records on over 1.8 million veterans who served in the Navy, enlisted between 1941 and 1945, and died prior to 2013. Each observation is identified by an individual's Social Security number, but name and location at death are not included. However, the date of enlistment into the Navy (a field also reported in muster rolls) is included, making this data particularly useful for linking in tandem with other data sources.

Further information on each data source, reliability, and variables contained is reported in Appendix B.1.

3.5 Sample for Main Analysis

Subsequent analysis focuses on a specific subset of individuals: white men born between 1905 and 1928 who can be linked to Navy records and at least one post-war migration period. I include only those whose state of enlistment matches their pre-war state of residence.[33] This restriction eliminates ambiguity about whether post-war locations represent moves relative to pre-war or wartime residences. The resulting sample comprises 380,000 individuals linkable to either their 1950 location or place of death.

Two key considerations arise when assessing how men in the sample may differ from the general population: (1) how well the linked sample represents the population of white men in rank-and-file Navy ship positions, and (2) how this Navy population compares to the general cohort of white men born between 1905 and 1928. Figure A.6 addresses these issues by comparing the linked sample to all men in the Muster Rolls data (Panel A) and to same-age men in the 1940 Census (Panel B). Each panel in Figure A.6 compares three linked samples: those linked to 1940 (All Links), those linked to both 1940 and 1950 (1950 Links), and those linked to both 1940 and death records (Death Links). Coefficients represent the standardized difference between the linked sample and the full population.

The linked sample appears largely representative, with all coefficients within 0.2 standard deviations of the general population. Notable selection patterns include a slight under-representation of individuals from the South and over-representation from the Midwest in both comparisons. When compared to the 1940 Census, the sample shows a lower share

[32] FOIA 23-09430

[33] Pre-war state of residence is determined from either the 1940 Census or the state of Social Security number issuance for those issued between 1937 and enlistment. Around 15% of men in my sample have a different state of enlistment than pre-war state of residence.

of individuals from farm households, likely reflecting the agricultural exemption for Navy service. Additionally, the sample exhibits slight positive selection relative to the general U.S. population, with higher education, occupation scores, and geographic mobility, consistent with typical patterns in linked samples.

4 Impact of WW2 Ship Networks on Migration

This section estimates the impact of geographic networks formed on Navy ships on post-war migration decisions. I begin by describing the assignment process of sailors to ships and provide evidence for conditional random assignment. I then show how variation in the geographic mix of shipmates impacted state of residence in 1950 and by time of death.

4.1 Empirical Strategy

Estimating the causal impact of exposure to shipmates from different states on migration patterns requires ruling out systematic assignment of sailors to ships based on their likelihood of moving or preference for specific locations (Manski, 1993). It also necessitates separating out the role of shipmates from other impacts of the Navy that might impact migration such as dislocation, exposure to specific locations through service, and networks formed in other aspects of service.

To disentangle network effects from these other channels that might impact migration, it is therefore crucial to understand the institutional details of how the Navy assigned sailors to ships. This section describes how sailors were assigned to ships and demonstrates that, after accounting for transportation logistics, the geographic composition of shipmates was effectively random. This section concludes by discussing how other impacts of Navy service may interact with shipmate networks.

4.1.1 Assignment to Ships

Historical accounts indicate that newly trained rank-and-file personnel were highly substitutable during World War II.[34] These men, who comprised approximately 75% of the Navy

[34] From *Administration of Navy Department in World War II* (Furer, 1960): "The distribution of enlisted personnel was in some respects more difficult than the distribution of officers because the numbers were greater, but the clerical work in the Bureau was less because the movement of enlisted men was usually accomplished in drafts and did not require the writing of orders for each individual. Eventually the distribution of such personnel was further simplified by the Navy Classification Code system. Under this system a code number was assigned to every enlisted man in the Navy which stood for his background of vital statistics and education, rating, and skill; this code number, in effect, described the man and his qualifications. Each billet in the Navy also had a code number describing the skills and special qualifications needed by an individual

force throughout the war, were typically young and minimally trained. Advancements in Navy bureaucracy and early computing technology facilitated a systematic assignment process for these sailors relative to previous wars. Upon completion of basic training, men were assigned to ships primarily based on immediate operational needs rather than individual characteristics or preferences.

The Navy's assignment process, while broadly random, was shaped by three logistical and operational considerations. First, men were transported in groups to minimize transportation costs,increasing the likelihood of serving with others who enlisted at similar times and from the same region.[35] Second, changing Navy policies, particularly lowering the minimum draft age to 18 in July 1942 and ending volunteer enlistment in late 1942, altered the composition of new entrants over time.[36] Third, different ship types required varying skill mixes, with submarines, for example, having a higher proportion of skilled technical positions than destroyers.[37]

Given these factors, I consider assignment to Navy ships to be random when conditioning on three variables: state of enlistment, first quarter received on a ship, and ship type. Thus, what distinguishes two men from Pennsylvania who enlisted in early 1943 and served on destroyers is the chance assignment to a specific ship based on immediate needs at the time they completed basic training. Consequently, any difference in outcomes can be attributed to experiences in the Navy and not underlying differences in preferences or migration propensities.

It is likely that other experiences of Naval service impacted a person's migration likelihood beyond networks formed with shipmates. Certain forces were broadly shared across all servicemen, including dislocation, access to GI benefits, and the opportunity cost of service years. However, other factors that influence migration may be particular to a person's service experience: networks formed during training, experiences in other places along the ship's route, the impact of combat and trauma, promotions, etc. While these forces likely influence migration, I abstract away from them, by assuming ship networks are orthogonal

to fill it satisfactorily. Bringing the code number of the individual and of the billet description together resulted in locating the man qualified to fill the billet. This work was accomplished in the Bureau by the use of punched cards and tabulating machines, and removed the process from the realm of excessive detailed clerical work."

[35]The transportation of enlisted personnel was usually made in groups known as drafts. Orders were issued to individuals on the rare occasions when some highly qualified specialist was needed in an emergency to fill a specific billet. The drafts were normally distributed to ships through the organizations of the service commands or through receiving stations. At times a sudden requisition for a large draft made heavy demands on the distribution service of Bureau of Personnel (BuPers).

[36]Executive Order 9279 reduced the inflow of men from agricultural states and eliminated the option to choose Navy over Army service.

[37]Some ships also had specific physical requirements or relied on volunteers, as with the inherently dangerous submarine service.

to these other forces.[38]

4.1.2 Balance Tests

The identification strategy relies on the assumption that assignment to Navy ships was random, conditional on three factors: state of enlistment, first quarter on ship, and ship type. To validate this assumption, I test whether individual characteristics from the 1940 Census predict the composition of shipmates they encounter, after controlling for these three factors.

I estimate the following balance test to show the relationship between a sailor's pre-war characteristics and the ethnic, geographic, or economic composition of his later shipmates.

$$y_{ik} = \beta \underbrace{X_i}_{\text{Baseline Characteristics}} + \underbrace{\gamma_{g(i),h(i)}}_{\text{Ship type, quarter, and state FE}} + \varepsilon_{ik} \tag{1}$$

Here, y_{ik} represents various characteristics of the ship k on which individual i served, X_i are baseline characteristics of individual i, and $\gamma_{g(i),h(i)}$ includes fixed effects for enlistment state, first quarter on ship, and ship type.[39] Coefficients are scaled to represent 1/100 of a standard deviation of y_{ik}.

Figure 2 presents the results of these balance tests, with each panel showing balance on a different ship characteristic y_{ik}. Each plot displays three specifications: no controls, state fixed effects, and fixed effects for state, ship type, and first quarter on ship.[40]

Without controls, baseline characteristics and ship characteristics are correlated: individuals living in a different state than their birth state have, on average, a 2 p.p. (20% of a SD) greater share of shipmates from western states. This relationship is not surprising given the coordinated distribution of people to training and ships, and reflects that people from western states in 1940 are 31 p.p. more likely to be living in a different state than birth state relative to the national average.

Adding state controls almost fully eliminates any relationship between baseline and ship characteristics. For instance, conditional on being from the same state, a person living in a different state than their birth state is on average on ships with only a 0.1 p.p. (1% of a SD)

[38] One factor that may be of particular concern would be if individuals who are systematically exposed to people from a particular state also spend more time in that state during the course of service. Current work in progress is to establish groupings of ships that had similar geographic paths over the course of World War II.

[39] Throughout the paper, $h(i)$ will represent a person's pre-war residence, while $g(i)$ represent his "type" classified by when he served and what type of ship he served upon.

[40] Panel C which shows the share of co-ethnic shipmates also include controls for own ethnicity in all three specification.

greater share of shipmates from Western states. For almost all other ship characteristics and baseline characteristics, the coefficient is not statistically different from zero. Finally, adding time and ship type controls largely does not change coefficients. However, it importantly attenuates the coefficient on age such that it is not statistically different from zero, reflecting that people who enlist at similar times are more likely to be of similar ages.

Two main factors explain the residual correlation between baseline characteristics and ship assignment. First, state of enlistment imperfectly proxies the Navy's spatial distribution constraints, as enlistees from different parts of large states may systematically end up on different ships. Second, the bias towards similar shipmates scales with the number of same-day boarders, and individuals boarding with larger groups tend to be somewhat differently selected.[41] An alternative specification with county controls and restricting the sample to those with less than 10% same-day boarders fails to reject the joint nullity in a Wald test, supporting these explanations. In Section 4.3.3, I re-run core estimates using this more conservative specification and all results of the paper remain unchanged.

These results provide strong support for the assumption that, conditional on the key variables, assignment to ships was essentially random with respect to characteristics that might influence later migration decisions.

4.2 Main Empirical Specifications

I estimate the impact of ship networks on migration along two margins: out-migration and directed migration. Out-migration refers to the likelihood of leaving one's geographic origin due to exposure to shipmates from other areas. Directed migration captures how exposure to people from a specific location influences the probability of migrating to that location.

The analysis examines these effects at three levels of geographic granularity (state, Census division, and Census region) and two points in time (1950 and time of death).

The following equation represents the causal impact of shipmates on out-migration:

$$\underbrace{y_{ikmt}}_{\text{Moves out of own } m} = \beta_{mt} \underbrace{X_{ikm}}_{\text{Share of ship not from own } m} + \underbrace{\gamma_{h(i),g(i)}}_{\text{Ship type, quarter, and state FE}} + \varepsilon_{ikmt} \tag{2}$$

where y_{ikmt} indicates whether individual i from boat k has moved out of their original location m by time t. X_{ikm} is the share of shipmates not from location m. β_{mt} measures the effect of increased exposure to out-of-area shipmates on the probability of out-migration by period t.

[41] Looking across all sailors in the sample, shipmates who boarded on the same day account for 5% of total shipmate exposure at the median and 25% of exposure at the 75th percentile.

For directed migration, I estimate:

$$\underbrace{y_{ikjt}}_{\text{Lives in } j} = \beta_{jt} \underbrace{X_{ikj}}_{\text{Share of Ship from } j} + \underbrace{\gamma_{h(i),g(i)}}_{\text{Ship type, quarter, and state FE}} + \varepsilon_{ikjt} \tag{3}$$

Here, y_{ikjt} indicates whether individual i lives in location j at time t. X_{ikj} represents the share of shipmates on boat k from location j. β_{jt} captures the effect of increased exposure to shipmates from j on the probability of living in j at time t. $\gamma_{h(i),g(i)}$ includes fixed effects for enlistment state, enlistment time, and ship type.

These two migration margins provide complementary insights into how exposure to shipmates from other geographic locations influences an individual's migration decision. Networks affect migration decisions by altering the option value of moving to specific locations through the addition of network ties. The directed migration specification tests this mechanism directly by measuring how random exposure to shipmates from specific geographic locations changes the likelihood of moving to that location. The out-migration specification then considers the aggregation of these direct migration effects by measuring the impact of total exposure to individuals from other places on the decision to migrate at all.

While informative, this analytical framework does have limitations: each specification isolates only one component of the ship network and measures its impact on a single margin of the migration decision. To address these limitations and provide a more comprehensive analysis, Section 5 employs a discrete choice model that considers the role of the entire ship network on the full migration decision.

4.3 Results

This section presents empirical findings on how Navy ship networks influenced post-war migration patterns. I first examine the impact of exposure to shipmates from other places on out-migration. Then I analyze the role of exposure to shipmates in predicting whether people systematically move to where their shipmates live prior to the war (directed migration). Finally, I explore heterogeneity in these effects and run a series of robustness tests to validate the findings.

4.3.1 Impact of Navy Networks on Out-Migration

Figure 3 quantifies the relationship between exposure to shipmates from different geographic locations and post-war out-migration. Panel A demonstrates how out-migration rates vary with the share of shipmates from one's own location across states, Census divisions, and Census regions. Each sub-panel presents a bin scatter of out-migration by 1950 and over a person's lifetime, using within-home state quintiles to group observations. The plots reveal that men with a higher share of shipmates from their own state were substantially less likely to migrate - men in the highest quintile of own-state shipmates were 1.1 percentage points less likely to leave their state by 1950 compared to those in the lowest quintile, with this gap widening to 2 percentage points by time of death.

Panel B presents regression results from equation (2), which formalizes this relationship between shipmates and out-migration by providing causal estimates of how exposure to shipmates from different locations influences both short-term (by 1950) and long-term (by time of death) out-migration. Exposure to shipmates from different geographic areas significantly increases the likelihood of subsequent migration. A one standard deviation (6 p.p.) increase in exposure to out-of-state shipmates raises the likelihood of out-of-state migration by 0.5 percentage points in 1950 and 1.1 percentage points by time of death, corresponding to a 4.2 percent increase by 1950 and 2.8 percent increase by time of death. While the coefficients in Panel B of Figure 3 are smaller for out-of-division and out-of-region migration, this difference in effect size reflects that the rate of out-migration over longer distances is lower. Proportionally, a one standard deviation increase in exposure to out-of-division and out-of-region shipmates increases the likelihood of migration by 4.5% and 6.5% in 1950 and 4.4% and 4.1% by time of death, respectively.

The effect of exposure to shipmates from other places on out-migration exhibits striking proportionality across different geographic levels and time periods, despite substantial variation in baseline migration rates. For instance, out-of-state migration increases from 12% in 1950 to 42% by time of death, and out-of-region migration increases from 5%-26%. While the proportional network effect size on out-migration is somewhat weaker over a person's lifetime than by 1950, particularly for out-of-state migration, these relative magnitudes are more stable than those observed for directed migration. The relatively consistent proportional effect of networks on out-migration over time aligns with a scenario where direct ties drive initial moves, followed by second-order mobility effects stemming from network-induced migration.

4.3.2 Impact of Navy Networks on Directed Migration

This section examines how random exposure to shipmates on Navy ships affects the likelihood of a person migrating to the places where those shipmates lived prior to the war. I analyze the impacts of shipmate exposure on directed migration across three levels of geographic granularity: Census regions, Census divisions, and states.[42] Figure 4 presents the main

[42]There are nine Census divisions nested within the four Census regions (West, Midwest, Northeast, and South).

results at the Census division level. Panel A reports the causal estimates from equation (3), while Panel B shows these estimates normalized as percent increases in directed migration from a one standard deviation increase in exposure. Appendix Figure A.9 illustrates the underlying relationship between exposure and migration rates for each Census division, while Appendix Figures A.8 and A.7 present complementary results at the region and state levels.

Navy networks most strongly influence migration to areas rapidly growing in the post-war period. For instance, a one standard deviation (7.2 p.p.) increase in exposure to shipmates from Pacific states is associated with a 18 percent (0.4 p.p.) increase in migration to those states by 1950. As discussed in Section 2 the West was by far the fastest growing area of the country in the 1940s. Similar patterns extend to other high-growth areas, with significant network effects observed for migration to the coastal South (10% increase) and Mid-Atlantic Census divisions (10% increase).

While the magnitude of the effect of shipmate exposure on migration is largely proportional to the underlying migration rate to specific areas, the Midwest presents a notable exception to this pattern. Despite being the most common migration destination in 1950 for men in my sample and experiencing substantial in-migration from the South, Navy networks had little influence on migration to these states in the Midwest. This divergence likely stems from pre-existing migration patterns between the South and Midwest for white individuals; the established flow of people between these regions meant that potential migrants already had access to information and support networks, making new Navy connections less critical for these moves.

Looking across both time horizons, while the impact of increased shipmate exposure is more pronounced in absolute terms by time of death, the proportional effect is roughly two times as large in 1950. For instance, while a one standard deviation increase in exposure to shipmates from Pacific states increases the likelihood of moving by 0.4 p.p. in 1950 and 0.9 p.p by time of death, the proportional impact by time of death is only 55 percent of that in 1950. This pattern suggests that these direct ties were most salient in the short run, but continued to be influential over a person's lifetime.

Examining the results across levels of geographic granularity reveals that treatment effects of directed migration are unevenly distributed within regions and divisions. For example, a one standard deviation increase in exposure to people from Western states increases the likelihood of moving to the West by 17%. In contrast, a one standard deviation increase in exposure to Californians raises the likelihood of moving to California by 23%. State-level heterogeneity in treatment will be further explored in the discrete choice analysis in Section 5.

These results highlight two potential mechanisms through which Navy networks influenced migration: (1) by facilitating moves to high-growth, high-opportunity areas, and (2) by providing connections to areas where individuals were less likely to have pre-existing ties.

4.3.3 Heterogeneity and Robustness

Heterogeneity

This section examines whether Navy networks had differential effects on migration across subgroups of sailors. In particular, I explore whether individuals with geographically isolated networks are disproportionately influenced by exposure to new connections. Theoretically, individuals from rural areas, lower-income backgrounds, or those without prior migration experience might have fewer ties outside their home region and thus benefit more from access to geographically expanded networks.

Figure A.11 tests for these differential effects by estimating equation (2) separately for subgroups defined by baseline characteristics from the 1940 Census. Panel A presents the direct coefficients from these estimations, and Panel B converts these coefficients into standardized effects by showing the percent increase in migration probability from a one standard deviation increase in out-of-state exposure, facilitating comparison across groups with different baseline migration rates.

The results show remarkable consistency in the impact of Navy networks across subgroups. A one standard deviation increase in exposure to out-of-state shipmates raises migration probabilities by 4%-5% regardless of pre-war characteristics. This uniformity likely reflects the relatively homogeneous nature of the sample–predominantly young, white, non-college educated men.

While the overall effects are similar, point estimates are slightly larger for men from rural and lower-income counties, who likely had more limited pre-war networks. However, these differences are not statistically distinguishable from zero. The evidence thus provides only modest support for the idea that geographically isolated individuals benefit more from new network connections.

Robustness

In this section, I address a main threat to the causal interpretation of the results: systematic assignment of individuals to ships based on migration propensity. To address potential assignment bias, I conduct two exercises. First, I predict the likelihood of moving to Pacific states by 1950 using 1940 baseline characteristics. I construct this prediction for both the linked Navy sample and the full population of white men born between 1905 and 1928. Figure A.13 shows that predicted migration is uncorrelated with ship characteristics for both samples, indicating that baseline characteristics do not explain the observed increase in migration to Pacific states.

Second, I reconstruct my main estimates using a more restricted sample and robust controls. Figure A.14 presents results using a sample where individuals' exposure to those they boarded with is less than 10% of their total shipmate exposure. This restriction minimizes potential selection effects from boarding groups. I also include more granular geographic controls (county) and fully interacted fixed effects. The estimated treatment effects remain consistent across all sample and fixed effect combinations, supporting the robustness of the main results.

These exercises provide strong evidence that the observed migration patterns are not driven by systematic assignment bias or confounding factors, but rather reflect the causal impact of Navy-formed networks on migration decisions.

5 Discrete Choice Model of Migration with Networks

In this section, I introduce a discrete choice model of cross-state migration that incorporates networks from Navy ships. I document that the expansion of network ties increased the likelihood of migration, largely driven by direct connections over long distances. Then, I use model estimates to quantify the contribution of ship networks to aggregate migration trends.

5.1 Preference over Location

The following model describes a World War II veteran's migration decision in the post-war period. Following Blumenstock, Chi, and Tan (2023), this model embeds place-specific networks into a discrete choice framework.

The utility an individual i receives from moving to state d in period t:

$$U_{idt} = \underbrace{\mathbb{1}\left(d \neq h(i)\right)\left[\beta^{dest}_{h(i)dt} X_{k(i)d} + \pi_{h(i)dg(i)t}\right]}_{\text{Utility from not home state}} + \underbrace{\mathbb{1}(d = h(i))[\beta^{home}_{dt} X_{k(i)d} + \gamma_{dg(i)t}]}_{\text{Utility from home state}} + \underbrace{\varepsilon_{idt}}_{\text{T1EV Logit shock}} \tag{4}$$

Here, $k(i)$ indexes the Navy ship served on during the war, $g(i)$ represents a person's type (defined by the category of ship and first quarter served), and $h(i)$ denotes pre-war state of residence. Individuals choose the state that maximizes their utility. Assuming ε follows a type-I extreme-value distribution yields conditional logit preferences over states.

This model captures three key components of the migration decision: First, type-specific push-pull factors between states ($\pi_{h(i)dg(i)t}$ and $\gamma_{dg(i)t}$) such as distance, wages, and amenities. Second, network-specific factors ($\beta^{dest}_{h(i)dt} X_{k(i)d}$ and $\beta^{home}_{dt} X_{k(i)d}$) capture the impact of ship ties, with β^{home} and β^{dest} allowing for differential effects between the home state and other states.

The $\beta^{dest}_{h(i)dt}$ term further allows the value of networks in potential destinations to vary by both origin and destination state, reflecting that ties may have different values depending on the specific state pair involved in a potential move. Finally, all other factors are captured by the individual idiosyncratic term ε.

The discrete choice model is similar in specification to the OLS models discussed in Section 4. Both measure the role of the geographic mix of shipmates on the migration decision, but the functional form of the discrete choice models yields some significant advantages. First, the discrete choice model controls for multilateral resistance in ship networks– when measuring the impact of an additional shipmate from California on moving to California, the OLS estimation (Equation (3)) is agnostic to the characteristics of other shipmates, while the discrete choice model controls for the relative attractiveness of other network ties. Second, while the OLS estimates captures the average impact of additional ties to a location on subsequent migration to that place, the discrete choice model allows destination locations to vary in attractiveness based on pre-war residence. This heterogeneity in the discrete choice model occurs through two channels: the mean attractiveness of destination states varies by origin state and type (captured by $\pi_{h(i)dg(i)t}$), and the value of network ties to that destination state vary by origin state (captured by $\beta^{dest}_{h(i)dt}$).

5.1.1 Parameterizing Network Effects

The value of additional network ties may vary by home and destination state, reflecting the complex ways in which exposure to people from different places can influence migration decisions. These network effects encompass several mechanisms: information sharing, reduction of migration costs, job referrals, and the amenity value of friendship. These various channels suggest that the value of additional network ties vary both based on existing migration patterns and the relative desirability of destinations. I will not distinguish between these mechanisms but instead allow for enough flexibility to capture the general effect of networks on migration decisions.

Both information sharing and reduction of migration costs suggest that network ties are more valuable in places with sparse existing migration networks. New connections to shipmates from locations with fewer established ties could therefore be particularly influential in shaping migration decisions.

The relative desirability of destinations also plays a role in determining network value. If expectations about wages and amenities are anchored to personal experiences or national averages, network connections would have varying impacts based on destination desirability. Information about places with higher wages or better amenities than one's current location would increase the perceived value of moving. Conversely, learning about lower wages or fewer amenities elsewhere could decrease migration likelihood.

In addition to reducing migration costs and providing information, networks may also be valued as a place-based amenity. If people value having friends in a location, this would contribute equally to the appeal of all destinations, regardless of their economic characteristics.

Ideally, to capture these varied effects, one would estimate network effects for every origin and destination state combination. However, this estimation procedure would require a prohibitively large number of parameters given the available data. To capture key variations in network value while maintaining model parsimony, I parameterize network effects along two dimensions: destination region and pre-existing migration patterns:

$$\beta^{\text{dest}}_{h(i)dt} = \sum_{r \in \text{Reg}} \beta^{r,dest}_{t} \cdot \mathbb{1}(d \in r) + \beta^{\text{dist}}_{t} \log \text{dist}_{h(i)d} \tag{5}$$

Here "dist" is the distance in kilometers between the centroids of state h and d. The term $\beta^{r,dest}_{t}$ allows the value of new ties to vary across regions, capturing regional differences in information content and network density. The β^{dist}_{t} term proxies for existing networks between states, reflecting that the value of new networks may scale with the density of pre-existing ones.[43]

The home network effects capture ways in which an increase in the share of people from one's home state changes the utility of remaining in that state, encompassing both intra-state moves and improving the option value of local labor market opportunities. To account for regional variation in the value of expansion to one's home network, I parameterize home networks effects as:

$$\beta^{\text{home}}_{dt} = \sum_{r \in \text{Reg}} \beta^{r,home}_{t} \cdot \mathbb{1}(h \in r)$$

This formulation allows the value of home-state ties to vary across regions, capturing potential regional differences in the importance of local networks.

5.1.2 Estimation and Identification

The conditional choice probability that individual i chooses to live in state d by time t is:

$$P_{d|it} = \frac{\exp\left(V_{idt}\right)}{\sum_{d'} \left(\exp\left(V_{id't}\right)\right)} \tag{6}$$

[43] An alternative specification uses existing migration flows instead of distance: $\beta^{\text{dest}}_{h(i)dt} = \sum_{r \in \text{Reg}} \beta^{r,dest}_{t} \cdot \mathbb{1}(d \in r) + \beta^{\text{mig}}_{t} \text{cond mig}_{h(i)d,35-40}$, where cond $\text{mig}_{h(i)d,35-40}$ represents migration flows between states from the 1935-1940 census, capturing established migration patterns.

where V_{ikdt} is the common component of the indirect utility function.[44] The log-likelihood associated with this choice probability and parameters Γ_t can be written as

$$LL(\Gamma_t) = \sum_i \sum_d I_{idt} P_{d|it} \tag{7}$$

where I_{idt} is an indicator for if individual i lives in state d in time t. I estimate this model separately for both periods using Poisson Pseudo Maximum Likelihood (PPML). As described in Guimaraes, Figueirdo, and Woodward (2003) in the general case, and Sotelo (2019) for its use in gravity models, PPML is a tractable and numerically equivalent alternative to estimating discrete choice models via maximum likelihood estimation.[45]

The parameters of the model (β^{home}_{dt} and $\beta^{dest}_{h(i)dt}$) are identified if the idiosyncratic component of utility is independent of the characteristics of an individual's ship network $X_{k(i)d}$. As discussed in Section 4, assignment to ships is random conditional on type $g(i)$ and pre-war residence $h(i)$. Therefore with the inclusion of type-specific home and destination effects ($\pi_{h(i)dg(i)t}$ and $\gamma_{dg(i)t}$), this assumption is satisfied.

However, using a fully flexible parameterization $\pi_{h(i)dg(i)t}$ would result in over one hundred thousand distinct estimates. I therefore assume that type-specific (g) utility from living in state d is independent of state h.

$$\pi_{h(i)dg(i)t} = \bar{\pi}_{h(i)dt} + \delta_{dg(i)t}$$

This additional assumption allows for a tractable parameterization of push-pull factors between states.

5.2 Results

Figure 5 presents the results from estimating equation (7). Panel A reports estimates for migration by 1950 and by time of death, with destination network effects parameterized relative to distance as shown in equation (5). Panel B visualizes the implied pairwise coefficients for $\beta^{dest}_{h(i)dt}$ and β^{home}_{dt} for both periods. Since I control for origin-destination push-pull effects through $\pi_{h(i)dg(i)t}$ and $\gamma_{dg(i)t}$, these network effects can be interpreted as the causal impact

[44] $U_{ikdt} = V_{ikdt} + \varepsilon_{idt}$ where

$$V_{idt} = \underbrace{\mathbb{1}\,(d \neq h(i))\,[\beta^{dest}_{h(i)dt} X_{k(i)d} + \pi_{h(i)dg(i)t}]}_{\text{Utility from not home state}} + \underbrace{\mathbb{1}(d = h(i))[\beta^{home}_{dt} X_{k(i)d} + \gamma_{dg(i)t}]}_{\text{Utility from home state}}$$

[45] For other recent uses see Blumenstock, Chi, and Tan (2023) and Dingel and Tintelnot (2023). I implement the procedure using the Stata package `ppmlhdfe` (Correia, Guimarães, and Zylkin, 2020)

of Navy networks on the migration decision.

The results indicate that network ties to states outside one's home state positively influence migration decisions for most origin-destination state pairs. Specifically, $\beta^{dest}_{h(i)d}$ is positive for 78% of home-destination pairs in 1950, increasing to 92% when examining migration by time of death.

The positive value of β^{dist} suggests that additional network connections are more influential in migration decisions when a new connection is from a distant state, aligning with the interpretation that Navy networks are most valuable where pre-existing networks were sparse. An alternative specification supports this interpretation (Appendix Figure A.15), as using pre-existing migration patterns yields similar relative utility estimates, with correlations of 0.92 in 1950 and 0.83 by time of death.

The value of network ties varies significantly across destination regions. Notably, ties to the Midwest are least valuable, with all instances of negative utility from destination networks occurring in this region in 1950. This finding is consistent with results from Section 4.3.2, which showed that new ties to the Midwest did not increase migration to those states. Ties within one's origin state have minimal impact, suggesting that the out-migration results in Section 4.3.1 are driven by shifts in ties to other locations rather than differential effects of home networks.

5.3 Counterfactual Exercises

Using the estimated discrete choice model, I conduct several counterfactual exercises. First, to compare the predictions of the conditional logit model with the OLS results from Section 4.3, I estimate the predicted change in migration to California with a ten percentage point increase in exposure to Californian shipmates. Second, to extend the analysis from Section 4.3, relative to examining how exposure to shipmates from a single location impacts migration, I measure how the overall geographic mix of shipmates affects levels of out-migration and directed migration. Third, I conduct a back-of-the-envelope calculation of the share of population growth in California between 1935 and 1940 that can be explained by World War II networks.

5.3.1 Counterfactual: Moving to California

Using California as a case study, I compare migration predictions from the discrete choice model against OLS estimates. California provides an appealing test case as the fastest growing state between 1940 and 1950 in population. For both models, I estimate the impact of a ten percentage point increase in the share of Californian shipmates on the probability of moving to California.

For the discrete choice model, I construct state-specific predictions by comparing two counterfactual ships for each origin state h: a low-exposure ship $\underline{k}_h$ and a high-exposure ship $\bar{k}_h$. These ships differ in their share of Californian shipmates by ten percentage points, with all other state shares held proportionally constant at their empirically observed levels.

$$\Delta_{10} P_{\text{CA}|ht} = \frac{1}{\sum_{i \in h}} \sum_{i \in h} \left(P_{\text{CA}|i\bar{k}_{h(i)}t} - P_{\text{CA}|i\underline{k}_{h(i)}t} \right) \tag{8}$$

where $\Delta_{10} P_{\text{CA}|ht}$ represents the change in probability of moving to California by period t for individuals from state h, averaged across all individuals i, given a ten percentage point increase in Californian shipmates. To get a comparable estimate using OLS, I estimate equation (3) for directed migration to California, which yields a uniform effect across all origin states.

Figure 6 illustrates how Navy networks differentially impact migration to California across origin states. Panel A maps overall migration flows to California for men in my sample, Panel B shows the percentage point increase in migration from a ten percentage point rise in Californian shipmates, and Panel C presents these effects as percent change relative to average flows.[46] Figure A.16 replicates this analysis for Texas as the destination state.

While the OLS and discrete choice model predict similar average effects, the discrete choice model reveals substantial geographic heterogeneity in migration responses to Californian shipmates. A ten percentage point increase in Californian shipmates raises average migration to California by 0.5 percentage points in 1950 and one percentage point by death in both models. However, the discrete choice model shows that effects vary from 0.2 to 2.0 percentage points in 1950 and 0.3 to 1.9 percentage points by death, with the largest impacts concentrated in states neighboring California. These heterogeneous effects mirror overall migration patterns, indicating network effects amplify existing migration propensities– consistent with results from Section 4.3 showing the impact of migration networks is largely proportional to average migration rates.

Navy networks most strongly influenced migration to California for individuals from states with limited pre-existing migration ties to California. Converting these effects into percent changes reveals that exposure to Californian shipmates increased migration from Eastern states by 30-38% compared to just 7-20% for Western states in 1950, with similar but attenuated patterns by death (18-22% for Eastern states versus 7-11% for Western states). These differential effects by state of origin demonstrate how Navy networks attenu-

[46] Percentage changes in Panel C are computed by dividing the values in Panel B divided by Panel A, and then multiplying by 100 for each state.

ated gravity patterns in migration flows, with network ties offsetting distance frictions where average migration flows are weakest.

5.3.2 Counterfactual: Out-Migration and Directed Migration

To better understand how Navy networks shaped migration decisions, I decompose migration into two components: the probability of moving out-of-state and the probability of choosing a specific destination state conditional on moving. This analysis extends Section 4 by considering how a ship's entire geographic composition, rather than individual components, affected migration patterns. It also allows me to explore heterogeneity in how the impact of ship networks varied across different origin and destination states.

I conduct two counterfactual exercises that compare observed migration patterns to those predicted if individuals had been assigned to ships with network composition matching the average for others from their home state. The first exercise quantifies how variation in shipmate characteristics affects the likelihood of leaving one's home state. The second analyzes how networks influence the choice of destination state, conditional on moving.

Out-Migration Counterfactual

For each home state h, I measure out-migration relative to the average ship.[47] The counterfactual exercise is expressed as:

$$\Delta P_{d \neq h(i)|it} = \left(P_{(d \neq h(i))|k(i)it} - P_{(d \neq h(i))|\tilde{k}_{h(i)}it} \right) \tag{9}$$

where k_h is the average ship for home state h, and $\Delta P_{d \neq h|ht}$ captures the change in out-migration probability compared to placement on the average ship. This exercise extends Section 4.3.1 by incorporating the full ship composition rather than just the share of out-of-state shipmates. The counterfactual accounts for both the presence of home-state connections and the relative attractiveness of connections to different destinations. For each origin state, I present results in terms of a standard deviation of $\Delta P_{d \neq h|ht}$ across individuals. This represents how a one standard deviation change in effective variation of shipmate characteristics affects the probability of moving out-of-state, measured in percentage points.

Ship networks substantially influence an individuals' propensity to leave their home state, with effects varying greatly across states of origin. Panel B of Figure 7 shows a one standard deviation change in ship characteristics raises out-migration probabilities by 0.2 to 1.1 percentage points in 1950 and 0.1 to 1.6 percentage points by time of death. When scaled by average migration rates (Panel C), these changes represent increases of 1.3% to 6.9% in

[47] The average ship for state h is computed as the geometric mean of the vector of ship shares across all individuals from state h.

1950 and 0.8% to 3.9% by death, aligning with the average effect of share of shipmates from different states on out-migration of 4.2% (1950) and 2.8% (death) found in Section 4.3. The impact on longer-distance moves is slightly larger: Appendix Figures A.17 and A.18 show that similar network variation increases cross-division and cross-region migration by between 1.3% and 10.7% by 1950. The large heterogeneity in the impact of Navy networks on out-migration is only weakly correlated with the average out-of-state migration rates shown in Panel A (correlation of 0.58 in 1950 and 0.02 by time of death).

Geographic patterns in where networks were mostly influential in driving migration shifts over time. In 1950, networks generated the largest increases in state-level out-migration from the Sun Belt, Upper Midwest, and the Northeast, while by time of death the strongest effects were concentrated in the Northeast and Pacific. For longer-distance moves, however, the pattern was more stable: Figure A.18 shows that networks consistently had the largest impact on cross-region migration out of the Midwest and Northeast across both time horizons.

Conditional Directed Migration Counterfactual

This counterfactual examines how networks influenced destination choice conditional on moving out-of-state, and is computed as:

$$\Delta \text{Dir Mig}_{idt} = \frac{P_{(d|ik(i)t)}}{1 - P_{(h(i)|ik(i)t)}} - \frac{P_{(d|i\tilde{k}_{h(i)}t)}}{1 - P_{(h(i)|i\tilde{k}_{h(i)}t)}} \tag{10}$$

where k_h is the average ship composition for individuals from home state h. One standard deviation of variation in $\Delta\text{Dir Mig}_{idt}$ across individuals from state h measures the effective impact (in percentage points) of shipmate characteristics on migrating to state d, conditional on leaving one's own state.

Figure 8 Panels A and B illustrate how increased exposure to shipmates from specific states shapes bilateral migration flows. A one standard deviation increase in exposure raises conditional directed migration between certain state pairs by up to 4 percentage points in 1950 and 1 percentage point by time of death, with the largest effect observed for migration from Colorado to California. These treatment effects ($\text{SD}_h\Delta\text{Dir Mig}_{idt}$) are strongly correlated with average bilateral migration flows between states (correlation of 0.75 in both periods). Converting the treatment effects into percent changes, as shown in Panel B, reveals that variation in shipmate exposure increases conditional migration between state pairs by 5-15% in most cases, with some pairs experiencing increases in conditional migration of up to 50%.

While network effects on migration flows are largely proportional to average conditional migration patterns, these effects are particularly strong for moves to distant, higher-income states. Comparing average conditional migration flows and bilateral network effects with differences in state income levels in 1940 reveals this systematic pattern: average conditional migration flows show little correlation with income differences between states, while migration driven by ship network variation exhibits a strong positive correlation with state income differences, indicating these networks particularly facilitated moves to higher-opportunity areas. Networks effects are also stronger for geographically distant state pairs, suggesting these connections helped overcome barriers to long-distance migration.

Figure 8 Panel C aggregates these bilateral effects to measure how variation in shipmate exposure shapes migration flows to each destination state. For each destination state, the impact of a one standard deviation increase in exposure is calculated by first computing the percent change in migration from each origin state, then taking a weighted average across origin states, with weights proportional to migration flows. This normalized measure captures how increased Navy exposure affects the likelihood of choosing a particular destination state over alternatives.

The strongest effects emerge for migration to rapidly growing states in the West and South. California shows the largest response, with a 17% increase in migration following a one standard deviation increase in exposure to Californian shipmates. Similarly large effects appear for Texas (11%) and Florida (12%). These states also received the highest share of overall conditional migration flows - Florida and California each account for 15% of conditional migration, followed by Texas at 5%. This pattern suggests that Navy networks amplified existing migration trends toward high-growth areas, making already-attractive destinations even more likely to receive migration inflows.

5.3.3 Counterfactual: Share of population growth attributable to war networks

To contextualize the magnitude of these migration effects and their potential role in explaining the rise in overall migration during this period, I conduct a counterfactual exercise focusing on California's population growth between 1945 and 1950. California, the fastest-growing state during this period, experienced a population increase of 1.4 million (15% growth) over these five years.[48]

Using predictions from the discrete choice model, I estimate migration flows to California under two scenarios: one where everyone had average exposure to Californians (8%) and another with no exposure. I then compute the predicted increase in flows from each state of origin. To scale these predictions to the entire population of World War II veterans, I use the state-of-enlistment shares from my Navy data and Army enlistment data, combined with general population estimates for each branch (9.5 million Army, 3.5 million Navy).

[48] Population data sourced from https://fred.stlouisfed.org/series/CAPOP (Bureau, 2024)

This approach predicts 47,500 additional veterans moving to California due to World War II networks. Assuming these veterans moved with their households, and using the average household size in the sample of 4.1 (in 1950), this back-of-the-envelope calculation implies approximately 195,000 people moved to California due to wartime networks. This figure represents roughly 14% of California's total population growth during this period.

The back-of-the-envelope calculation likely underestimates the total effect for two reasons. First, it only considers migration spillovers at the household level, ignoring potential broader network effects once links between places are established. Second, this estimate is based solely on the differential cross-Navy ship effect, potentially missing a higher "base level" effect from other aspects of war experience or general Navy service. I note, however, this exercise does not account for potential crowd-out–in the absence of these war network-induced moves, alternative migration flows to California might have occurred.

I replicate this exercise for two other fast-growing states during this period: Texas and Florida. Texas experienced population growth of 930,000 (14%) between 1945 and 1950, while Florida's population increased by 330,000 (14%). Using the same methodology, I estimate that World War II networks account for 3.9% of Texas's growth and 2.8% of Florida's growth between 1945 and 1950. These findings suggest that wartime networks potentially played a significant role in shaping migration patterns, particularly for migration to the West.

6 Shared Ethnicity and Network Formation

This section investigates whether shared ethnicity among shipmates influences post-war migration patterns. The motivation for this analysis stems from the nature of the treatment in this study: exposure to individuals from diverse geographic backgrounds. While general exposure may influence migration choices, it is likely that closer ties formed during service would have a more pronounced effect on post-war decisions.

I focus primarily on shared ethnicity as a key factor influencing network formation and subsequent migration decisions. The salience of ethnic differences among white individuals by country of origin in the 1940s suggests that co-ethnic interactions may have been particularly influential in the context of racially homogeneous but ethnically diverse Navy ships.

This approach builds on previous research by Carrell, Sacerdote, and West (2013), which demonstrated that patterns of endogenous network formation with peer randomization can significantly influence outcomes. By analyzing the differential impact of shared ethnicity on later migration, I identify one characteristic through which men randomly placed together on a ship, non-randomly formed closer ties.

Appendix C.1 extends this analysis to three additional factors that may influence the strength of ties formed on ships: (1) occupational proximity on ships, (2) socioeconomic background, and (3) the level of combat the ship experienced during the war. These factors provide further insight into the nuanced ways in which shipboard experiences might have shaped post-war migration patterns.

6.1 Empirical Framework

To estimate the impact of the ethnic mix of shipmates on an individual's migration decision, I expand upon the empirical frameworks in both Section 4 and Section 5.

First, I augment equation (2) to examine how ethnic composition influences out-migration by 1950 or by death. The analysis includes three specifications. I begin by estimating the impact of the overall share of co-ethnic shipmates. I then decompose out-of-state shipmates into those who share the individual's ethnicity and those who do not. Finally, I incorporate the share of co-ethnic shipmates from one's home state to examine whether ethnic ties influence the decision to stay. For all specifications, I include state by ethnicity fixed effects in addition to the usual time and ship type fixed effects.

Next, I modify the utility specification within the discrete choice framework (Equations (4) and (5)) by decomposing the share of shipmates from each state who share the individual's ethnicity and those who do not.

$$\begin{aligned}
\beta_{hd}^{\text{dest}} X_{k(i)d} &= \beta_{\text{dest}}^{\text{co-eth}} X_{k(i)d}^{\text{co-eth}} + \beta_{\text{log dist}}^{\text{co-eth}} \log \text{dist}_{h(i)dt} X_{k(i)d}^{\text{co-eth}} + \beta_{\text{dest}}^{\neg\text{co-eth}} X_{k(i)d}^{\neg\text{co-eth}} + \beta_{\text{log dist}}^{\neg\text{co-eth}} \log \text{dist}_{h(i)dt} X_{k(i)d}^{\neg\text{co-eth}}, \\
\beta_{d}^{\text{home}} X_{k(i)d} &= \beta_{\text{home}}^{\text{co-eth}} X_{k(i)d}^{\text{co-eth}} + \beta_{\text{home}}^{\neg\text{co-eth}} X_{k(i)d}^{\neg\text{co-eth}}.
\end{aligned} \tag{11}$$

Here, $X_{ikdt}^{\text{co-eth}}$ represents the share of shipmates from state d who are also co-ethnics, while $X_{ikdt}^{\neg\text{co-eth}}$ denotes the share of shipmates from state d who are not co-ethnics. By decomposing effects into the contribution of co-ethnic shipmates and shipmates of a different ethnicity, I capture how the presence of co-ethnic shipmates differentially influences migration and whether the differential impact of co-ethnic shipmates varies with pre-existing network density (proxied by $\log \text{dist}$). This specification also includes ethnic-specific destination pull factors in the discrete-choice model to account for potential differences in spatial returns to migration across ethnic groups.

6.2 Results

The analysis shows that shared ethnicity between shipmates significantly influences the likelihood of migration, suggesting stronger social ties formed between co-ethnic shipmates. Figure 9 presents both OLS estimates (Panel A) and discrete choice results (Panel B). The

OLS estimates reveal that a higher overall share of co-ethnic shipmates increases out-of-state migration by 1950, though this effect dissipates when examining lifetime migration patterns. This increased mobility in 1950 operates through two channels: first, co-ethnic shipmates from other states have a stronger impact on migration compared to non-co-ethnic shipmates; second, a higher share of co-ethnic shipmates from one's home state also increases overall geographic mobility.[49]

The discrete choice model corroborates these findings, showing that a higher share of co-ethnic shipmates from any state increases the likelihood of out-of-state migration. This broader increase in mobility from co-ethnic exposure may reflect enhanced network centrality within ships–due to the presence of more co-ethnic shipmates, men form closer ties with shipmates of other ethnicites from different states.

Panel B allows for direct comparison between co-ethnic and different-ethnicity shipmate effects on migration probability. Appendix Figure A.19 shows the predicted coefficient for β_{hd}^{dest} for when shipmates are co-ethnics or of a different ethnicity correlated with the distance between states. The results demonstrate that co-ethnic shipmates had substantially stronger effects on migration. In 1950, a co-ethnic tie was on average 2.5 times more influential than a non-co-ethnic tie. While this effect moderated over time, co-ethnic ties remained on average 1.5 times more influential by time of death. As shown in Figure A.19, the relative importance of ethnic ties varies with distance. In 1950, co-ethnic ties had a largely uniform impact on the likelihood of moving across destinations regardless of distance, while the impact of non-co-ethnic ties increases with distance between states, suggesting the relative value of co-ethnic ties was highest for close moves.

These results indicate that shared ethnicity significantly shaped network formation aboard ships and influenced subsequent migration patterns. The presence of more co-ethnic shipmates not only strengthened specific geographic ties but also increased overall mobility, highlighting how social connections formed during service affected post-war migration decisions.

In Appendix C.1, I present results on the impact of on-ship occupational proximity and pre-war income of shipmates on migration propensity. I find little evidence that either factor meaningfully changes the likelihood of migration, though both factors present substantial measurement challenges. Finally, in Figure A.12, I test for heterogeneity in treatment effects across different ship sizes and levels of combat exposure. I find no measurable impact of these factors on the magnitude of the treatment effect.

[49] These two channels can be seen in columns 2 and 3 when the share of shipmates from out-of-state is decomposed into the share of shipmates who are from out-of-state and co-ethnic and the share of shipmates who are from out-of-state and of a different ethnicity.

7 Returns to Networked Migration

This section leverages random variation in Navy ship networks to estimate returns to migration. The previous analysis established that networks formed during Navy service influenced both whether individuals moved and their choice of destination after the war. An important question remains: did individuals induced to move through these Navy networks experience greater economic opportunity?

Using an instrument for the probability of moving derived from the discrete choice model in Section 5, I estimate both prime-age (1950) and lifetime (death) returns to migration. This approach provides a unique quasi-experimental setting for studying the impact of network-facilitated migration on later earnings.

7.1 Navy Network Migration Instrument

To estimate how migration causally affects earnings, I construct instruments predicting the probability of individuals moving out-of-state, out-of-region, and to the Pacific by 1950 and by time of death. These predicted probabilities are derived from conditional choice probabilities (equation (6)) using estimated parameters from the discrete choice model in Section 5.

The instrument for the predicted probability of moving depends on two components. First, observable fixed characteristics that inform ship assignment – an individual's pre-war residence $h(i)$ and their type $g(i)$ (defined by ship type and first quarter on ship) – determine common push-pull factors between states. Second random variation in geographic composition of shipmates determine the role of networks in driving migration. Therefore, this approach distills multi-dimensional variation in shipmate characteristics into a single instrument, where conditional on $h(i)$ and $g(i)$, all variation in the instrument is random. The instrument, therefore, captures how random chance in ship assignment influences a person's likelihood of later migration.

Using this instrument, I estimate the returns to migration using two-stage least squares regression.

$$\begin{aligned}\underbrace{y_{ikt}}_{\text{Lives in } y} &= \alpha \underbrace{P_{y|it}}_{\text{Probability of moving to } y} + \underbrace{\gamma_{h(i),g(i)}}_{\text{Ship type, year, geography FE}} + \epsilon_{ikt} \\ \log \text{Inc}_{ikt} &= \beta\widehat{y_{ikt}} + \gamma_{h(i),g(i)} + \nu_{ikt}\end{aligned} \tag{12}$$

where y_{ikt} represents whether individual i in time t is living in y, where y represents living in a different state, a different region, or a Pacific state.[50] In 1950, I proxy individual income using occupational score.[51] By time of death, I proxy income using the median household income in the zipcode a person is last known to reside.[52] To extend this exercise, I also measure the impact of migration on non-pecuniary outcomes such as family formation and mortality (Appendix Table A.5).

In this setting, compliers are individuals who are marginal movers whose migration decision is sensitive to the characteristics of shipmates. An individual's classification as a complier, always-taker, or never-taker is governed by the push-pull factors between states as well as their idiosyncratic logit draws as described in Section 5. The idiosyncratic logit draw captures all residual factors besides ship networks and common state push-pull factors that influence a person's migration decision.[53]

Across states, the population of compliers, always-takers, and never-takers differs. For a high-growth state like California where pull factors to California often outweigh idiosyncratic reasons for moving, there is a high share of never takers–people who will always stay in California. Conversely, for a low-growth state like West Virginia where push factors are large, there is a greater share of always takers–people who move regardless of their ship experience. The instrument accounts for these differences: states with high shares of always takers and never takers will exhibit smaller variability in the instrument compared to states with a higher share of compliers. Therefore, for the IV estimate to be interpreted as the local average treatment effect (LATE) for compliers, the key assumption is that the variance in idiosyncratic error terms is constant across states (Angrist and Imbens, 1994; Nakamura, Sigurdsson, and Steinsson, 2022).[54]

Beyond these considerations, the standard instrumental variable assumptions apply. The relevance of the instrument holds if random variation in ship networks induces out-migration, which is demonstrated in the first stage results shown in Appendix Table A.21. The exogeneity of the instrument to baseline characteristics is shown in Appendix Table A.20. Monotonicity in the instrument holds if there are no defiers. In this context, defiers are individuals who behave contrary to the average effect. While the instrument allows for exposure

[50]For Pacific state moves, individuals originally from Pacific states are excluded.

[51]This is for two reasons. Income is sample-line in the 1950 Census and only available for 20% of individuals. Additionally, known irregularities in the current public release of the 1950 Census make occupational score a more reliable measure of income

[52]Median houshold income is reported in the year 2000 in nominal levels.

[53]A non-exhaustive list of residual factors that are captured in the idiosyncratic logit draw are: pre-existing network connections to other states, idiosyncratic preferences over amenities, idiosyncratic labor components that change spatial return to different states.

[54]For the LATE interpretation to hold under strictest assumptions, I need the first stage to be common across strata, particular pre-war residence. By controlling for push-pull factors directly in the instrument, only variation in the idiosyncratic logit term remains.

to shipmates from a particular state to increase or decrease the probability of moving, for monotoncity the directional effect of exposure to people from a destination state d must be stable for all people from the same origin state h.[55] Given this construction, I realistically assume the population of defiers to be negligible. Finally, the exclusion restriction states the geographic mix of shipmates must only impact incomes through migration. While this assumption is not directly testable, the main potential confounder would be if other peer characteristics correlated with geography drive increases in income through other channels.

7.2 Results

Network-facilitated migration led to large increases in lifetime earnings, as shown in Table 1, which reports estimates from equation (12). Coefficients for results in 1950 are reported in Appendix Table A.4. Columns 1-3 report the OLS estimates, while columns 4-6 report the IV estimates. The IV estimates show out-of-state movers experience 46 log points higher income, out-of-region movers see a 53 log point increase, and movers to the Pacific experience 69 log points increase, translating to 59%, 70%, and 99% higher earnings, respectively. These estimates indicate that men induced to move due to experiences on their Navy ship experienced much higher earnings over the course of their lifetimes. In 1950, I note small but significant returns to migration (1% for out-of-state moves, 3% for out-of-region moves, and 4% for moves to Pacific states).

The large pecuniary returns to migration raise the question of why migration rates were not higher in the absence of these network connections. As discussed in Nakamura, Sigurdsson, and Steinsson (2022), high returns to migration are consistent with substantial migration costs and young individuals being mismatched to their birthplaces. Networks formed during Navy service likely reduced these costs, facilitating moves to high-opportunity areas. Furthermore, if networks act as a place-based amenity, the Navy-formed connections may have lessened the trade-off between economic opportunity and network-based amenities.[56] Both of these factors may explain why individuals were willing to move when provided with new network connections, despite previously forgoing apparently lucrative migration opportunities.

The IV estimates being larger than the OLS estimates is perhaps surprising, especially as more geographically mobile people tend to be positively selected. However, several mech-

[55] i.e. It does not violate monotonicity if exposure to people from West Virginia decreases the likelihood of someone from Oregon moving out-of-state. It does violate monotonicity if for some people from Oregon exposure to people from West Virginia increases the likelihood of moving out-of-state and for others it decreases the likelihood of moving out-of-state

[56] Networks may act as a place-based amenity if individuals value having friends and community near where they live.

anisms could explain this pattern. First, compliers – those whose migration decisions are influenced by Navy networks – may experience larger returns than always-takers. If compliers live in areas with more geographically isolated networks, and these are also lower opportunity, then might reasonably be expected to have a higher return to migration than someone who is an always taker. Second, the IV specifically identifies returns to network-facilitated migration. While the OLS sample includes all Navy veterans, those who leverage networks to migrate may achieve better outcomes, perhaps through accessing higher-paying jobs through their connections. Third, given findings about the relevance of networks for long-distance moves in Section 5, the IV might be in part capturing substitution from some otherwise shorter distance cross-state moves to higher opportunity long-distance moves.

Appendix Figure A.24 illustrates the heterogeneity in returns to networked migration across various subgroups. The results reveal a distinct pattern: in the short run, higher-income, urban individuals from manufacturing households in 1940 experience the greatest benefits from migration. However, this pattern reverses in the long run. Individuals from lower-income, rural areas, and farm households in 1940 exhibit the highest lifetime returns to migration. This shift suggests that those who ultimately benefited most from migration took longer to realize these gains.

Table A.5 presents estimates for marriage, household size, and mortality, which allow me to examine the impact of migration on non-pecuniary outcomes. Migration shows no significant effect on marriage status, household size, or age at death. However, there is a strong positive impact on the likelihood of having a spouse born in a different state: moving out-of-state increases the probability of marrying someone born in a different state by 29 percentage points, conditional on being married.

While, these findings suggest that Navy networks were highly beneficial for those induced to move, I do note a few caveats. Median household income of zip codes is in nominal levels, so these returns may be in part offset by cost-of-living differences across places, and this effect might be particularly salient for the returns to moving to Pacific states. In addition, I am using zip code of residence at time of death as a proxy for income. Additionally, if there were additional peer effects from shipmates from high opportunity states, then this treatment might be a mixture of the returns to migration and the returns to other benefits such as human capital accumulation. The bundled treatment effect might in part explain the gap between the OLS and IV estimates.

The large returns to migration I am estimating for Navy sailors are largely in line with similar estimates from the literature.[57] For lifetime estimates, Nakamura, Sigurdsson, and Steinsson (2022) find that young individuals displaced by a volcanic eruption had 82% higher

[57] A comparison of estimates from the literature is shown in Appendix Figure A.23

earnings 35 years later. For a similar time period and looking to internal migration within the US, both Boustan (2016) and Collins and Wanamaker (2014) find large returns for migration from the South to the Midwest during the Great Migration, with returns for Black men exceeding 80%; Boustan (2016) finds returns for white men of around 60%. Ward (2022) examines brothers in early 20th century census records and finds that those who migrate out-of-state or out-of-region have 15% and 18% higher occupational scores by their 30s compared to their non-migrant siblings.

This analysis demonstrates that network-facilitated migration, as induced by Navy ship assignments during World War II, led to substantial increases in lifetime earnings. These findings highlight the importance of networks in facilitating beneficial migration and suggest that reducing migration costs through network formation can have significant long-term economic impacts.

8 Conclusion

This paper leverages the unique historical context of World War II Navy service to examine how the expansion of geographic networks influences migration patterns and economic outcomes. By constructing a novel dataset of Navy sailors during World War II and exploiting conditional random assignment to ships, the study provides causal evidence on the impact of newly formed social connections on geographic mobility. The findings demonstrate that exposure to individuals from diverse geographic backgrounds significantly increases both the likelihood of moving anywhere and in particular the choice of where to migrate.

The analysis reveals that war-formed networks may have played a crucial role in shaping post-war migration trends across the United States. Navy connections were particularly influential in driving migration from slower-growing areas to rapidly expanding regions particularly in the West, accelerating broader demographic shifts. These findings highlight how large-scale events such as wartime mobilization, can have lasting impacts on the geographic distribution of population and economic activity.

Importantly, the paper bolsters knowledge on the substantial economic benefits from migration. Individuals induced to move due to their expanded geographic networks relocated to areas where they achieved higher lifetime earnings. This finding underscores the potential for social connections to enhance economic opportunity and reduce spatial inequality in opportunity by facilitating access to more productive labor markets.

The results have broader implications for understanding social connectivity and economic mobility in the United States. The dramatic expansion of geographic networks during World War II likely contributed to higher levels of interstate connectivity persisting for years if not decades after the war. Given the important role of networks in facilitating migration and the potential for positive spillovers once disparate locations are connections, policies that foster connections between Americans from different areas may have long-lasting impacts on economic opportunity and spatial economic disparities.

References

Abramitzky, Ran, Leah Boustan, and Dylan Shane Connor. 2024. "Leaving the enclave: Historical evidence on immigrant mobility from the industrial removal office." *The Journal of Economic History*, 84(2): 352–394. Publisher: Cambridge University Press.

Abramitzky, Ran, Leah Boustan, and Katherine Eriksson. 2020. "Do immigrants assimilate more slowly today than in the past?" *American Economic Review: Insights*, 2(1): 125–141. Publisher: American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Abramitzky, Ran, Leah Boustan, Katherine Eriksson, James Feigenbaum, and Santiago Pérez. 2021. "Automated Linking of Historical Data." *Journal of Economic Literature*, 59(3): 865–918.

Acemoglu, Daron, David H. Autor, and David Lyle. 2004. "Women, War, and Wages: The Effect of Female Labor Supply on the Wage Structure at Midcentury." *Journal of Political Economy*, 112(3): 497–551. Publisher: The University of Chicago Press.

Aizer, Anna, Ryan Boone, Adriana Lleras-Muney, and Jonathan Vogel. 2020. "Discrimination and Racial Disparities in Labor Market Outcomes: Evidence from WWII."

Althoff, Lukas, and Christiane Szerman. 2024. "The G.I. Bill and Black-White Wealth Disparities."

Angrist, Joshua D., and Guido W. Imbens. 1994. "Identification and Estimation of Local Average Treatment Effect." *Econometrica*, , (2): 467–475.

Arellano-Bover, Jaime. 2022. "Displacement, Diversity, and Mobility: Career Impacts of Japanese American Internment." *The Journal of Economic History*, 82(1): 126–174.

Bailey, Amy Kate. 2011. "Race, Place, and Veteran Status: Migration among Black and White Men, 1940–2000." *Population Research and Policy Review*, 30(5): 701–728.

Barnhardt, Sharon, Erica Field, and Rohini Pande. 2017. "Moving to Opportunity or Isolation? Network Effects of a Randomized Housing Lottery in Urban India." *American Economic Journal: Applied Economics*, 9(1): 1–32.

Beaman, Lori A. 2012. "Social Networks and the Dynamics of Labour Market Outcomes: Evidence from Refugees Resettled in the U.S." *The Review of Economic Studies*, 79(1): 128–161.

Bedard, Kelly, and Olivier Deschênes. 2006. "The Long-Term Impact of Military Service on Health: Evidence from World War II and Korean War Veterans." *American Economic Review*, 96(1): 176–194.

Beine, Michel, Frédéric Docquier, and Çağlar Özden. 2011. "Diasporas." *Journal of Development Economics*, 95(1): 30–41.

Bergman, Peter, Raj Chetty, Stefanie DeLuca, Nathaniel Hendren, Lawrence F. Katz, and Christopher Palmer. 2024. "Creating Moves to Opportunity: Experimental Evidence on Barriers to Neighborhood Choice." *American Economic Review*, 114(5): 1281–1337.

Blumenstock, Joshua E., Guanghua Chi, and Xu Tan. 2023. "Migration and the value of social networks." *Review of Economic Studies*, rdad113. Publisher: Oxford University Press US.

Bound, John, and Sarah Turner. 2002. "Going to War and Going to College: Did World War II and the G.I. Bill Increase Educational Attainment for Returning Veterans?" *Journal of Labor Economics*, 20(4): 784–815.

Boustan, Leah Platt. 2016. *Competition in the promised land: Black migrants in northern cities and labor markets.* Vol. 39, Princeton University Press.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin. 2020. "Peer Effects in Networks: A Survey." *Annual Review of Economics*, 12(Volume 12, 2020): 603–629. Publisher: Annual Reviews.

Breen, Casey F., Maria Osborne, and Joshua R. Goldstein. 2023. "CenSoc: Public Linked Administrative Mortality Records for Individual-level Research." *Scientific Data*, 10(1): 802. Publisher: Nature Publishing Group UK London.

Bureau, U.S. Census. 2024. "Resident Population in California [CAPOP]."

Büchel, Konstantin, Maximilian V. Ehrlich, Diego Puga, and Elisabet Viladecans-Marsal. 2020. "Calling from the outside: The role of networks in residential mobility." *Journal of urban economics*, 119: 103277. Publisher: Elsevier.

Carrell, Scott E., Bruce I. Sacerdote, and James E. West. 2013. "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation." *Econometrica*, 81(3): 855–882. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA10168.

Carrington, William J., Enrica Detragiache, and Tara Vishwanath. 1996. "Migration with Endogenous Moving Costs." *The American Economic Review*, 86(4): 909–930. Publisher: American Economic Association.

Chetty, Raj, Matthew O. Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B. Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, and Matthew Jacob. 2022. "Social capital I: measurement and associations with economic mobility." *Nature*, 608(7921): 108–121. Publisher: Nature Publishing Group UK London.

Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." *American Economic Review*, 106(4): 855–902.

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

Clampet-Lundquist, Susan, and Douglas S. Massey. 2008. "Neighborhood Effects on Economic Self-Sufficiency: A Reconsideration of the Moving to Opportunity Experiment." *American Journal of Sociology*, 114(1): 107–143. Publisher: The University of Chicago Press.

Collins, William J., and Ariell Zimran. 2024. "Who Benefited from World War II Service and the GI Bill? New Evidence on Heterogeneous Effects for US Veterans."

Collins, William J., and Marianne H. Wanamaker. 2014. "Selection and economic gains in the great migration of African Americans: new evidence from linked census data." *American Economic Journal: Applied Economics*, 6(1): 220–252. Publisher: American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203-2425.

Correia, Sergio, Paulo Guimarães, and Tom Zylkin. 2020. "Fast Poisson estimation with high-dimensional fixed effects." *The Stata Journal: Promoting communications on statistics and Stata*, 20(1): 95–115.

Costa, Dora L., and Matthew E. Kahn. 2003. "Cowards and heroes: Group loyalty in the American Civil War." *The Quarterly Journal of Economics*, 118(2): 519–548. Publisher: MIT Press.

Costa, Dora L., Matthew E. Kahn, Christopher Roudiez, and Sven Wilson. 2018. "Persistent social networks: Civil war veterans who fought together co-locate in later life." *Regional Science and Urban Economics*, 70: 289–299.

Deryugina, Tatyana, Laura Kawano, and Steven Levitt. 2018. "The economic impact of Hurricane Katrina on its victims: Evidence from individual tax returns." *American Economic*

Journal: Applied Economics, 10(2): 202–233. Publisher: American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203-2425.

Dingel, Jonathan I, and Felix Tintelnot. 2023. "Spatial Economics for Granular Settings."

Dolfin, Sarah, and Garance Genicot. 2010. "What Do Networks Do? The Role of Networks on Migration and "Coyote" Use." *Review of Development Economics*, 14(2): 343–359. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9361.2010.00557.x.

Einiö, Elias. 2019. "Mixing the rich and poor: The impact of peers on education and earnings." *VATT Institute for Economic Research Working Papers*, 128.

Enamorado, Ted. 2021. "A Primer on Probabilistic Record Linkage." In *Handbook of Computational Social Science, Volume 2.* Routledge. Num Pages: 13.

Fellegi, Ivan P., and Alan B. Sunter. 1969. "A Theory for Record Linkage." *Journal of the American Statistical Association*, 64(328): 1183–1210.

Fetter, Daniel K. 2013. "How Do Mortgage Subsidies Affect Home Ownership? Evidence from the Mid-Century GI Bills." *American Economic Journal: Economic Policy*, 5(2): 111–147.

Fishback, Price, and Joseph Cullen. 2013. "Second World War spending and local economic activity in US counties, 1939–58 - Fishback - 2013 - The Economic History Review - Wiley Online Library."

Fishback, Price V., and Taylor Jaworski. 2016. "World War II and US Economic Performance." In *Economic History of Warfare and State Formation.* , ed. Jari Eloranta, Eric Golson, Andrei Markevich and Nikolaus Wolf, 221–241. Singapore:Springer.

Furer, Julius Augustus. 1960. *Administration of the Navy Department in World War II.* US Government Printing Office.

Garin, Andrew, and Jonathan L. Rothbaum. 2024. "The Long-Run Impacts of Public Industrial Investment on Local Development and Economic Mobility: Evidence from World War II."

Granovetter, Mark S. 1973. "The Strength of Weak Ties." *American Journal of Sociology*, 78(6): 1360–1380.

Guimaraes, Paulo, Octávio Figueirdo, and Douglas Woodward. 2003. "A tractable approach to the firm location decision problem." *Review of Economics and Statistics*, 85(1): 201–204.

Publisher: MIT Press 238 Main St., Suite 500, Cambridge, MA 02142-1046, USA journals ....

Guo, Yuchen, Matthew O. Jackson, and Ruixue Jia. 2024. "Comrades and Cause: Peer Influence on West Point Cadets' Civil War Allegiances."

Hall, Patricia Kelly. 2009. "Privileged moves: migration, race and veteran status in post-World War II America."

Harriman, W. Averell. 1948. "Internal Migration in the United States: April 1940 to April 1947."

Hausser, P. M., and A. J. Jaffe. 1947. "The extent of the housing shortage." *Law & Contemp. Probs.*, 12: 3. Publisher: HeinOnline.

Hendren, Nathaniel, Ben Sprung-Keyser, and Sonya R. Porter. 2022. "„The Radius of Economic Opportunity: Evidence from Migration and Local Labor Markets "." *CES Working Paper, Nr. 22–27*. Publisher: CES Working Papers.

Johnson, Marilynn S. 1994. *The second gold rush: Oakland and the East Bay in World War II.* Univ of California Press.

Kinnan, Cynthia, Shing-Yi Wang, and Yongxiang Wang. 2018. "Access to Migration for Rural Households." *American Economic Journal: Applied Economics*, 10(4): 79–119.

Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica*, 75(1): 83–119.

Koenen, Martin, and Drew Johnston. 2024. "Social Ties and Residential Choice: Micro Evidence and Macro Implications."

Ludwig, Jens, Greg J. Duncan, Lisa A. Gennetian, Lawrence F. Katz, Ronald C. Kessler, Jeffrey R. Kling, and Lisa Sanbonmatsu. 2013. "Long-Term Neighborhood Effects on Low-Income Families: Evidence from Moving to Opportunity." *American Economic Review*, 103(3): 226–231.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *The Review of Economic Studies*, 60(3): 531–542.

McKenzie, David, and Hillel Rapoport. 2010. "Self-selection patterns in Mexico-US migration: the role of migration networks." *the Review of Economics and Statistics*, 92(4): 811–821. Publisher: The MIT Press.

Michelman, Valerie, Joseph Price, and Seth D Zimmerman. 2022. "Old Boys' Clubs and Upward Mobility Among the Educational Elite*." *The Quarterly Journal of Economics*, 137(2): 845–909.

Molloy, Raven, Christopher L. Smith, and Abigail Wozniak. 2011. "Internal Migration in the United States." *Journal of Economic Perspectives*, 25(3): 173–196.

Moretti, Enrico, Claudia Steinwender, and John Van Reenen. 2023. "The Intellectual Spoils of War? Defense R&D, Productivity, and International Spillovers." *The Review of Economics and Statistics*, 1–46.

Morison, Samuel Eliot. 1963. *The Two-Ocean war: A short history of the United States navy in the second world war.* Little, Brown Boston.

Munshi, Kaivan. 2003. "Networks in the Modern Economy: Mexican Migrants in the U. S. Labor Market*." *The Quarterly Journal of Economics*, 118(2): 549–599.

Munshi, Kaivan. 2020. "Social Networks and Migration." *Annual Review of Economics*, 12(1): 503–524.

Nakamura, Emi, Jósef Sigurdsson, and Jón Steinsson. 2022. "The gift of moving: Intergenerational consequences of a mobility shock." *The Review of Economic Studies*, 89(3): 1557–1592. Publisher: Oxford University Press.

Porcher, Charly. 2020. "Migration with costly information." *Unpublished Manuscript*, 1(3).

Porcher, Charly, Eduardo Morales, and Thomas Fujiwara. 2024. "Measuring Information Frictions in Migration Decisions: A Revealed-Preference Approach."

Rosenbloom, Joshua L, and William A Sundstrom. 2004. "The decline and rise of interstate migration in the United States: Evidence from the IPUMS, 1850–1990." In *Research in Economic History*. Vol. 22 of *Research in Economic History*, 289–325. Emerald Group Publishing Limited.

Sacerdote, Bruce. 2001. "Peer effects with random assignment: Results for Dartmouth roommates." *The Quarterly journal of economics*, 116(2): 681–704. Publisher: MIT Press.

Sarvimäki, Matti, Roope Uusitalo, and Markus Jäntti. 2022. "Habit Formation and the Misallocation of Labor: Evidence from Forced Migrations." *Journal of the European Economic Association*, 20(6): 2497–2539.

Shue, Kelly. 2013. "Executive Networks and Firm Policies: Evidence from the Random Assignment of MBA Peers." *The Review of Financial Studies*, 26(6): 1401–1442.

Sotelo, Sebastian. 2019. "Practical aspects of implementing the multinomial pml estimator." *Ann Arbor: University of Michigan, mimeo.*

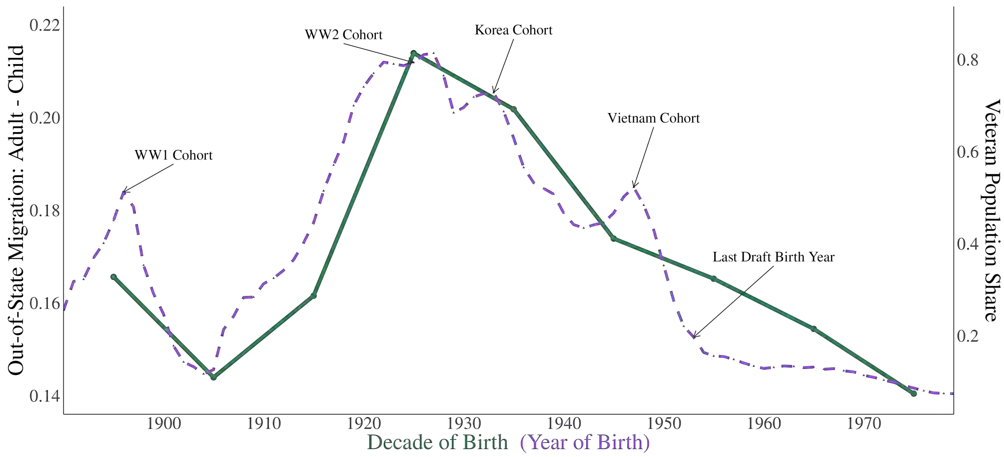
Stuart, Bryan A., and Evan J. Taylor. 2021. "Migration networks and location decisions: Evidence from US mass migration." *American Economic Journal: Applied Economics*, 13(3): 134–175. Publisher: American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203-2425.

Suandi, Matthew. 2022. "Essays in Economic History and Development Economics." PhD diss. UC Berkeley.

Ward, Zachary. 2022. "Internal Migration, Education, and Intergenerational Mobility: Evidence from American History." *Journal of Human Resources*, 57(6): 1981–2011.

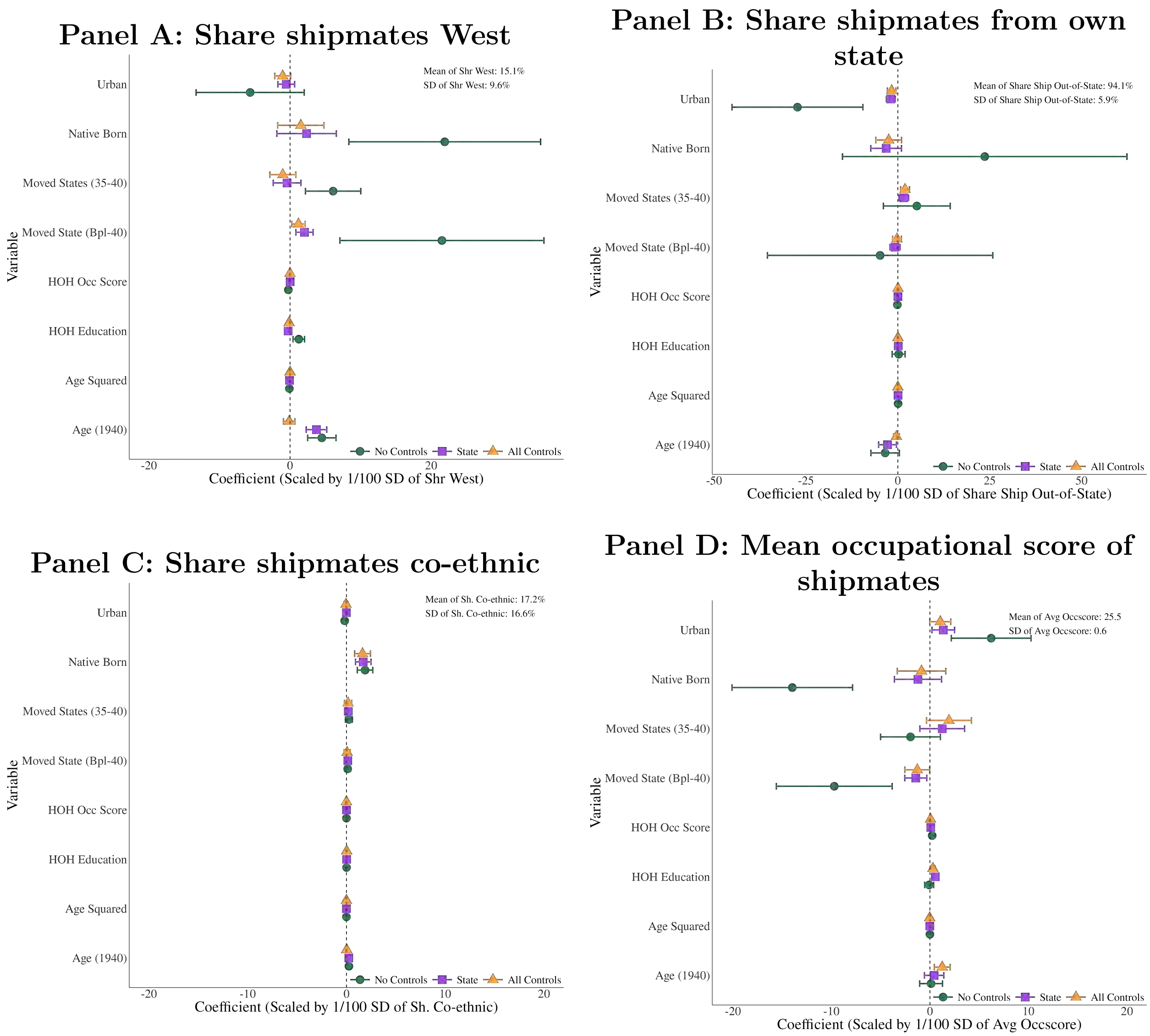
9 Exhibits

Figure 1: Change in probability of living outside state of birth from ages 10-19 to 30-39



Notes: This figure displays interstate mobility and military participation rates for white men born between 1890 and 1979. The green solid line shows the difference in probability of living outside one's birth state between ages 30-39 and 10-19, by birth decade. Levels are shown in Appendix Figure A.1. The dashed purple line represents the share of white men over the age of 35 who report being veterans, by birth year. Data source: Decennial census public use samples (1900-2010) and American Community Survey (2005-2024).

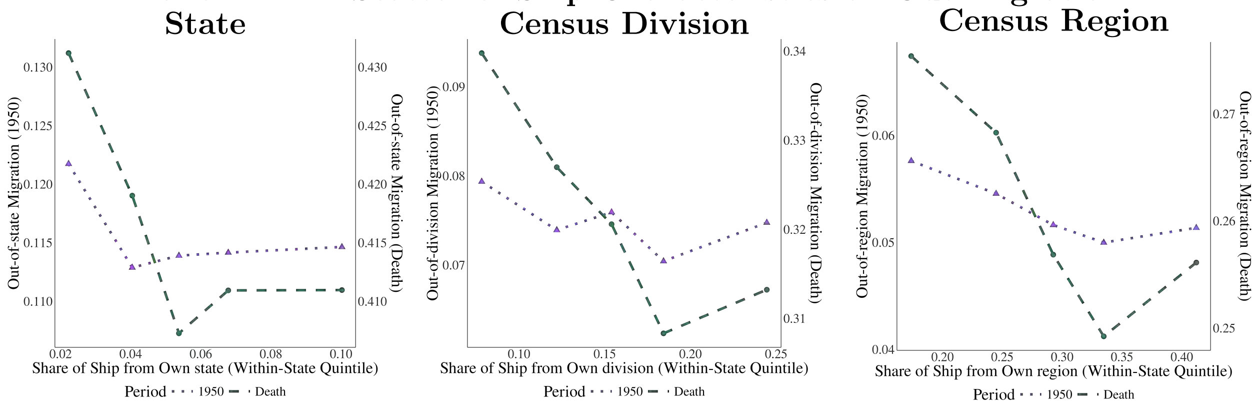
Figure 2: Balance Coefficient Plots



NOTES: This figure presents tests of whether baseline individual characteristics predict shipmate composition, controlling for state of enlistment, enlistment timing, and ship type. Each panel shows results from estimating equation (1) with a different shipmate characteristic as the dependent variable: Panel A shows the share of shipmates from Western states, Panel B shows the share from one's own state, Panel C shows the share of co-ethnic shipmates, and Panel D shows mean occupational score of shipmates. For each panel, coefficients from three specifications are displayed: no controls (green circles), state fixed effects only (purple squares), and the full set of fixed effects for state, ship type, and enlistment timing (yellow triangles). Panel C additionally controls for own ethnicity across all specifications. All coefficients are scaled to represent 1/100 of a standard deviation of the dependent variable. Horizontal bars represent 95% confidence intervals. Standard errors are clustered at the ship level.

Figure 3: Impact of Navy Networks on Out-Migration

Panel A: Bin Scatter of Ship Characteristics on Out-Migration



Panel B: Causal Impact of Shipmates on Out-Migration

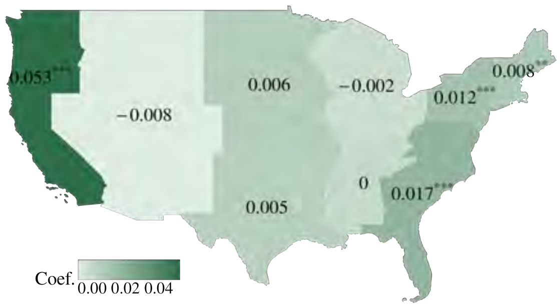
	1950			Death		
	State	Division	Region	State	Division	Region
	(1)	(2)	(3)	(4)	(5)	(6)
Sh. Ship Out-of-State	0.08***			0.19***		
	(0.02)			(0.03)		
Sh. Ship Out-of-Division		0.04***			0.17***	
		(0.01)			(0.01)	
Sh. Ship Out-of-Region			0.03***			0.11***
			(0.007)			(0.01)
Dep Var Mean	0.115	0.075	0.053	0.416	0.322	0.261
Observations	175,276	175,276	175,276	243,108	243,062	243,062
R^2	0.0288	0.0210	0.0143	0.0333	0.0356	0.0466
Within R^2	0.00013	0.00011	0.00020	0.00029	0.00062	0.00054
Ship type FE	✓	✓	✓	✓	✓	✓
State enlist FE	✓	✓	✓	✓	✓	✓
First quarter FE	✓	✓	✓	✓	✓	✓

Notes: Panel A shows the relationship between out-migration at three geographic levels (state, Census division, Census region) and shipmate characteristics. Each geographic level is represented by a separate plot, displaying migration by 1950 (purple, left axis) and by death year (green, right axis). Both axes span an equal range but are re-leveled to accommodate different baseline migration rates in each period. For each state of origin, five within-state quintiles are created based on the share of shipmates from the same state, division, or region. The average out-migration rate is calculated for each quintile, with a sample population-weighted average across home states reported on the plot. Panel B presents results from Equation (2) for the three geographic levels. Columns 1-3 report migration estimates from the 1950 Full Count census. Columns 4-6 report migration estimates over lifetime. All specifications include fixed effects for type of ship served on, state of enlistment, and the first quarter a person served on the ship. Standard errors are clustered at the ship-level.

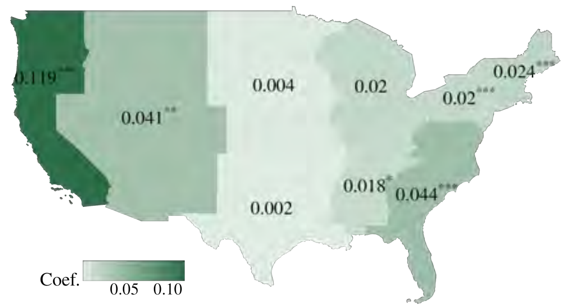
Figure 4: Impact of Navy Networks on Directed Migration (Census Divisions)

Panel A: Map of Directed Migration (p.p.)

1950

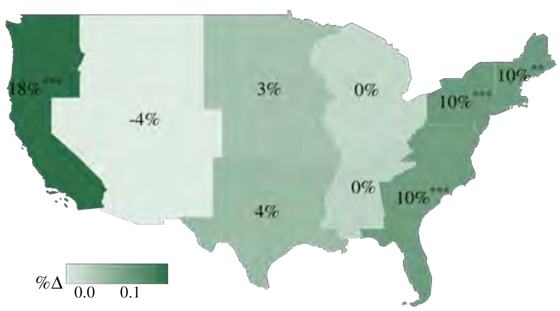


Death

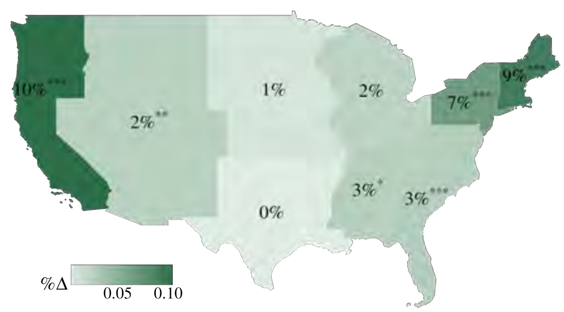


Panel B: Map of Directed Migration (% Change)

1950



Death



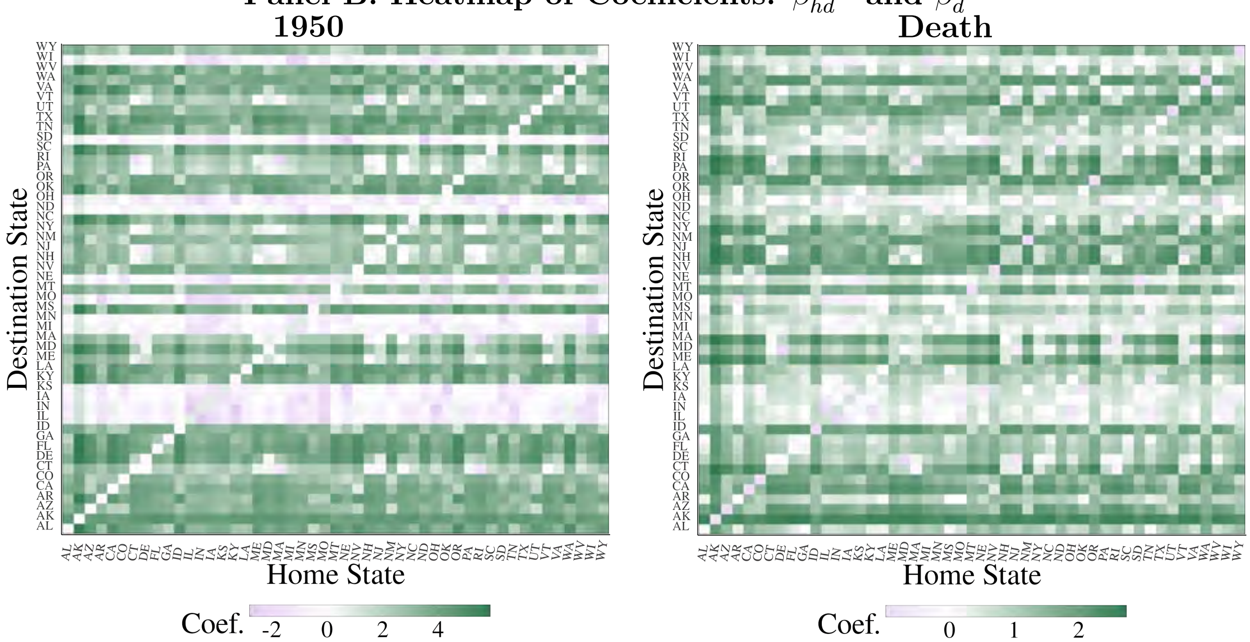
NOTES: This figure shows the impact of Navy ship networks on directed migration across Census divisions, by showing estimates from equation (3). Panel A presents β_{jt} showing the effect of increased exposure to shipmates from each Census division on migration to that division, measured both in 1950 and at time of death. The coefficients are depicted geographically on maps of the United States to highlight spatial patterns. Panel B shows these effects normalized as percentage increases in directed migration resulting from a one standard deviation increase in exposure to shipmates from each destination division. Standard errors are clustered at the ship-level.

Figure 5: Discrete Choice Model Coefficients

Panel A: Discrete Choice Estimates

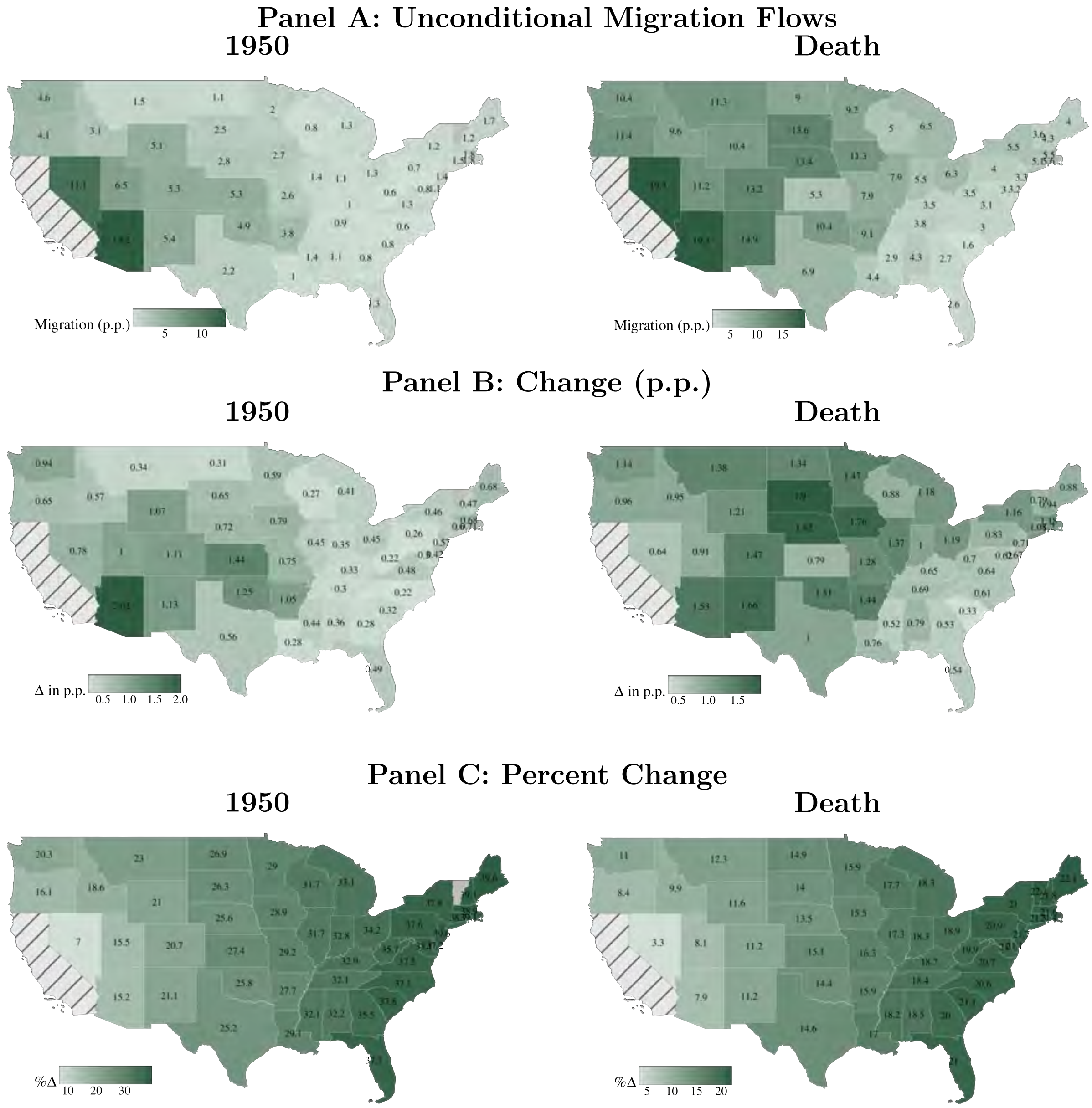
	Coefficient Estimates								
	β^{dist}	$\beta^{dest}_{Midwest}$	β^{dest}_{North}	β^{dest}_{South}	β^{dest}_{West}	$\beta^{home}_{Midwest}$	β^{home}_{North}	β^{home}_{South}	β^{home}_{West}
1950	1.35	-10.20	-7.85	-5.92	-7.05	0.79	0.18	-0.46	-0.23
	(0.34)	(2.30)	(2.21)	(2.38)	(2.59)	(0.44)	(0.30)	(0.54)	(0.41)
Death	0.77	-5.17	-3.90	-4.57	-4.10	0.59	0.69	-0.12	-0.91
	(0.18)	(1.25)	(1.16)	(1.28)	(1.37)	(0.23)	(0.13)	(0.26)	(0.20)
R-squared	0.65 (1950), 0.40 (Death)								
N	7,124,979 (1950), 11,276,968 (Death)								

Panel B: Heatmap of Coefficients: β^{dest}_{hd} and β^{home}_{d}



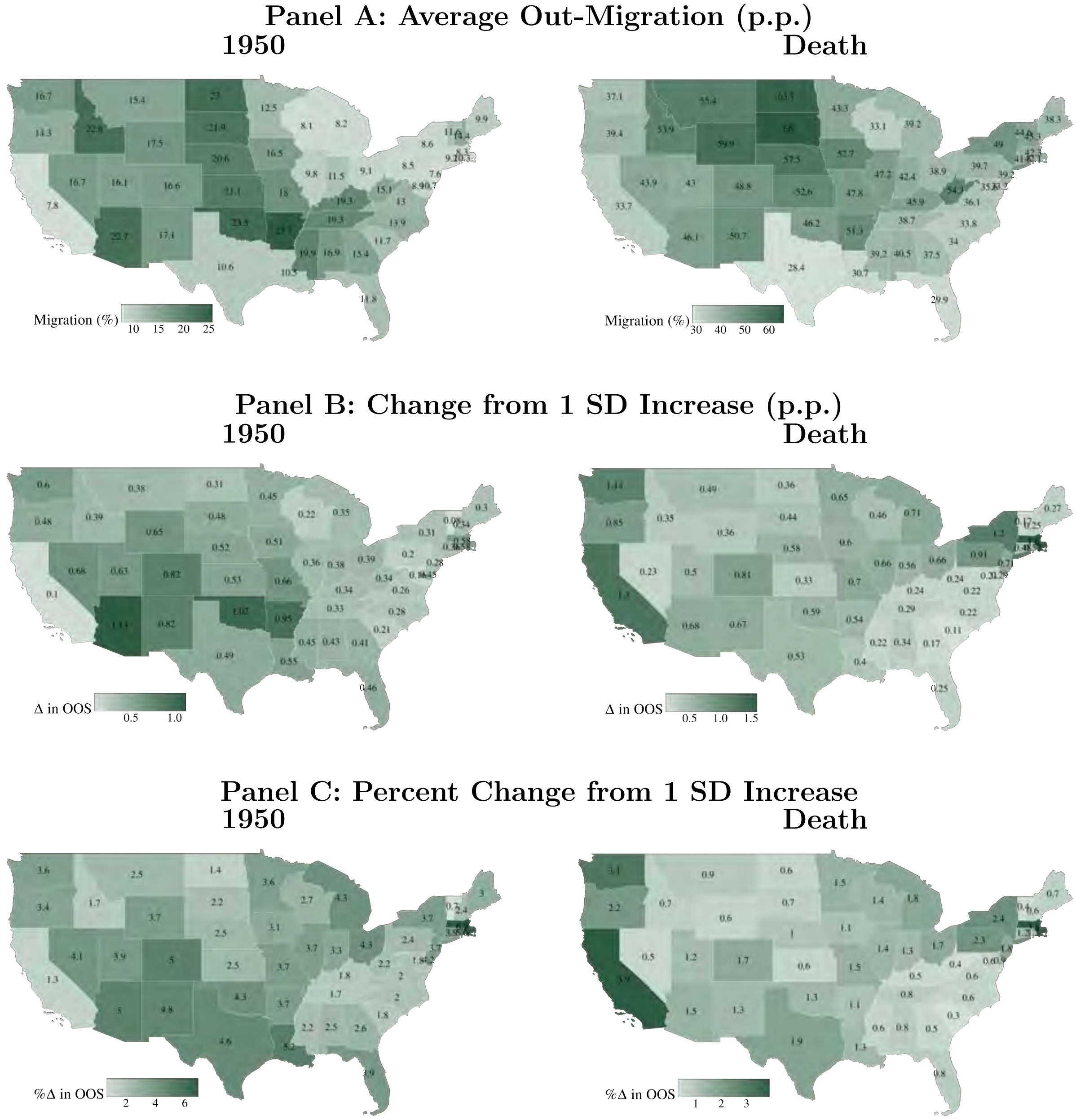
NOTES: Panel A reports discrete choice estimates from Equation (7). Row 1 reports estimates for migration by 1950, and Row 2 reports estimates for migrate by time of death. The pseudo-R^2 and number of observations is reported for each specification at the bottom. Coefficients are estimated by Poisson-Pseudo Maximum Likelihood (PPML) estimation. Panel B reports the implied home-destination coefficients of β^{dest}_{hd} and β^{home}_{d} from the coefficients estimated in Panel A and the parameterization described in Equation (5). Each cell is a home-destination state pair. Distance is computed as kilometers between state centroids. Logged distance between contiguous states vary between 4.1 (RI to MA) to 8.4 (ME to CA). The median origin-destination pair has a logged distance of 7.3 (MA to WI).

Figure 6: Impact of a 10 p.p. increase in Californian ship share on migration, by state of origin



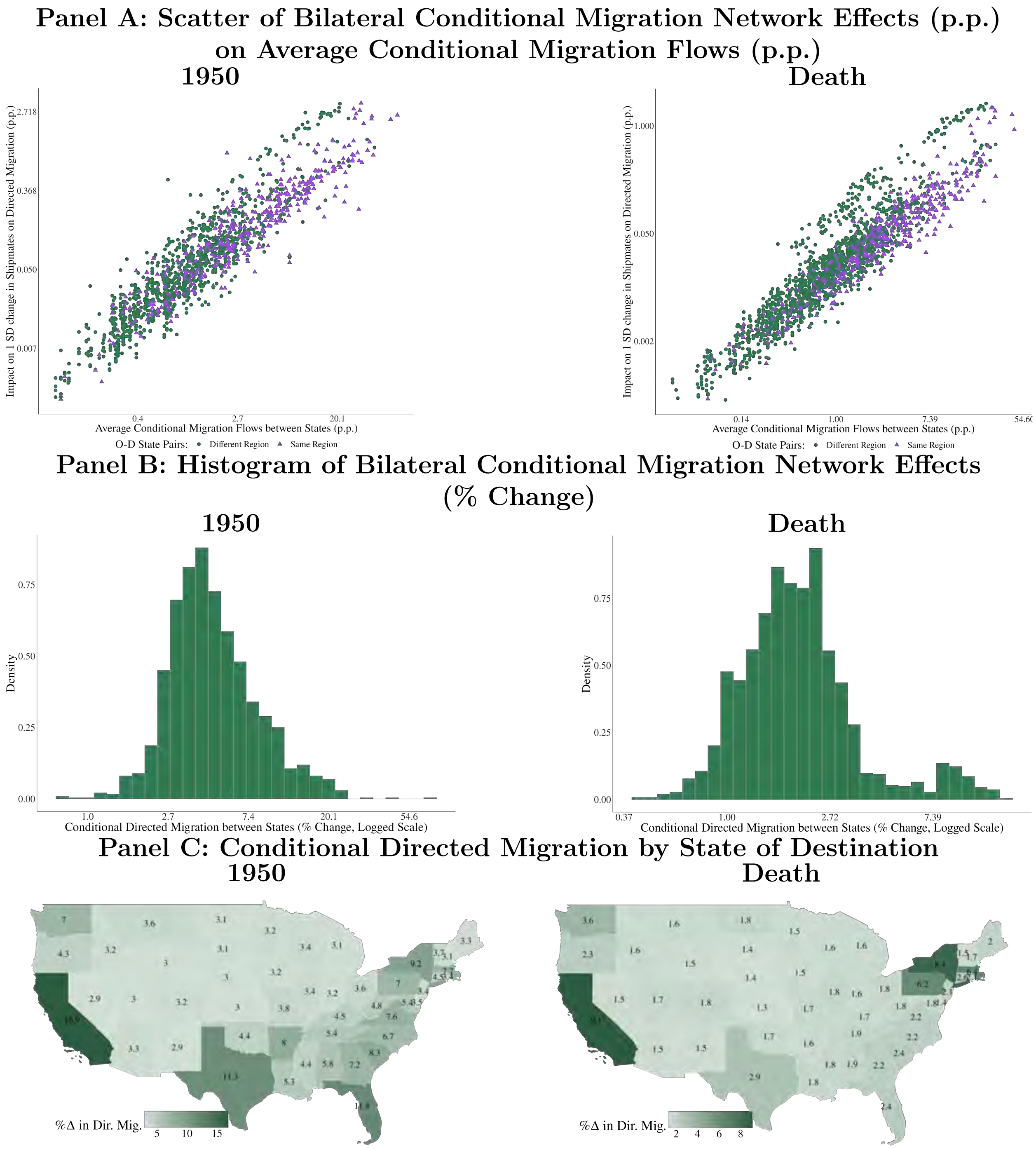
Notes: This figure presents results from the counterfactual exercising showing the impact of increased exposure to Californian shipmates on migration to California from each origin state. The counterfactual compares the predicted migration response from serving on a high-exposure ship (10 percentage points higher) to a low-exposure ship as described in Equation (8) using discrete choice estimates reported in Figure 5. Panel A reports average unconditional migration flows to California by origin state by 1950 and by a person's death. Panel B shows the increase in migration probability of serving on a high-exposure ship relative to a low-exposure ship in percentage points. Panel C displays the percent increase in migration probability relative to the average unconditional migration rate. The left map in each panel reports results in 1950, while the right map in each panel reports results by death.

Figure 7: Impact of one SD increase in ship exposure on cross-state migration, by state of origin



NOTES: This figure shows how variation in shipmate composition affects cross-state migration probabilities. Following Equation (9) and using estimates from Figure 5, the counterfactual, $(\Delta P_{d\neq h(i)|it})$, computes the difference in probability of moving out-of-state when assigned to one's actual ship versus a ship with average composition for their pre-war residence $h(i)$. Panel A displays average out-of-state migration rates in 1950 and by time of death for men in the linked sample. Panel B reports one standard deviation in $\Delta P_{d\neq h(i)|it}$ across individuals from state h, representing the change in migration probability from a one standard deviation increase in effective shipmate variation. Panel C shows this effect as a percentage change relative to average out-migration rates from each state (Panel B/Panel A). Each panel presents results for both 1950 (left) and time of death (right).

Figure 8: Impact of Navy shipmates on conditional migration



NOTES: This figure shows how variation in shipmate composition affects directed conditional migration between states. Following equation (10), I compute for each origin-destination state pair (h, d) the change in probability of moving to state d conditional on moving, comparing assignment to one's actual ship versus a ship with average composition for home state h (ΔDir Mig). Panel A presents scatter plots relating one standard deviation of ΔDir Mig (representing the impact of a one standard deviation increase in effective shipmate variation) to average conditional migration flows between state pairs. Panel B shows the distribution of these one standard deviation effects when normalized as percent increases over average conditional flows. Panel C aggregates these bilateral effects to the destination state level by computing weighted averages of the percent changes across origin states, with weights proportional to migration flows. Each panel compares effects in 1950 (left) to those measured at time of death (right).

Figure 9: Network Formation: Role of Co-ethnic Exposure in Migration

Panel A: OLS Estimates

	State Mover					
		1950			Death	
	(1)	(2)	(3)	(4)	(5)	(6)
Share Ship Out-of-State	0.08***			0.18***		
	(0.02)			(0.02)		
Sh. Co-ethnic	0.03***			-0.03		
	(0.01)			(0.03)		
Share Co-ethnic + Not State		0.10***	0.11***		0.15***	0.15***
		(0.02)	(0.03)		(0.008)	(0.04)
Share Not Co-eth + Not State		0.07***	0.08**		0.18***	0.19***
		(0.02)	(0.03)		(0.02)	(0.03)
Share Co-ethnic + Own State			0.05			0.03
			(0.05)			(0.12)
Observations	165,145	165,145	165,145	229,026	229,026	229,026
R^2	0.03703	0.03703	0.03703	0.04032	0.04032	0.04032
Within R^2	0.00013	0.00012	0.00013	0.00025	0.00025	0.00025
State by Ethnicity FE	✓	✓	✓	✓	✓	✓
Ship type FE	✓	✓	✓	✓	✓	✓
First Quarter FE	✓	✓	✓	✓	✓	✓

Panel B: Discrete Choice Estimates

	$\beta^{\text{co-eth}}_{\text{dest}}$	$\beta^{\text{co-eth}}_{\text{log dist}}$	$\beta^{\neg\text{co-eth}}_{\text{dest}}$	$\beta^{\neg\text{co-eth}}_{\text{log dist}}$	$\beta^{\text{co-eth}}_{\text{home}}$	$\beta^{\neg\text{co-eth}}_{\text{home}}$	R^2	N
1950	4.82	-0.11	-13.25	2.11	-1.94	0.62	0.65	5,243,710
	(4.64)	(0.67)	(2.42)	(0.34)	(0.59)	(0.29)		
Death	-5.28	1.01	-4.26	0.77	-1.34	0.67	0.41	8,017,913
	(2.67)	(0.37)	(1.21)	(0.17)	(0.31)	(0.13)		

NOTES: This figure presents estimates of how co-ethnic ties formed during Navy service influenced post-war migration patterns. Panel A reports OLS estimates for whether an individual moves out-of-state by 1950 (Columns 1-3) or by time of their death (Columns 4-6). All specifications include fixed effects for state of enlistment by ethnicity, first quarter on ship, and ship type, and standard errors are clustered at the ship level. Columns 1 and 4 include regressors for the share of shipmates from out-of-state and share of co-ethnic shipmates. Columns 2 and 5 include regressors for share of shipmates who are from out-of-state and co-ethnic, and those who are from out-of-state and not co-ethnic. Columns 3 and 6 add an additional regressor for the share of shipmates who are co-ethnic and from own's own state. Panel B reports discrete choice estimates from equation (11). Row 1 reports estimates for migration by 1950, and Row 2 reports estimates for migrate by time of death, and the pseudo-R^2 and number of observations is reported for each period t. Coefficients are estimated by Poisson-Pseudo Maximum Likelihood (PPML) estimation.

Table 1: Lifetimes Returns to Networked Migration

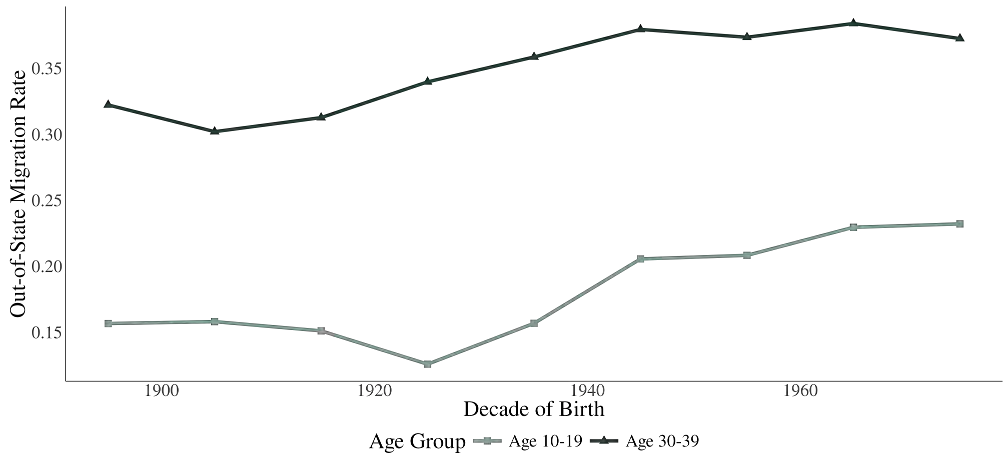
	Zipcode Inc Death (logged)					
	OLS			IV		
	State	Region	Pacific	State	Region	Pacific
	(1)	(2)	(3)	(4)	(5)	(6)
State Mover	0.05***			0.46***		
	(0.003)			(0.08)		
Region Mover		0.08***			0.53***	
		(0.004)			(0.09)	
Pacific Mover			0.61***			0.69***
			(0.006)			(0.13)
Observations	145,723	145,692	128,688	144,203	144,171	127,270
F-stat				228.67	201.38	103.16
1940 State by County FE	✓	✓	✓	✓	✓	✓
Ship type FE	✓	✓	✓	✓	✓	✓
First Quarter FE	✓	✓	✓	✓	✓	✓

NOTES: This table reports coefficients from Equation (12) showing the returns to migration by time of death (lifetime). Columns 1-3 report OLS estimates, while Columns 4-6 report IV estimates. In all specifications, the outcome is logged median household income of the last zip code the person lived in prior to their death. Income is reported for the year 2000 in nominal dollars. State mover is an indicator for if a person moved across state lines. Region mover is an indicator if a person moved between Census regions. Pacific mover is an indicator if someone who was not previously living in the Pacific Census division moved to a state in that division (excluding Alaska and Hawaii). In columns 4-6, the instrument is constructed as the predicted probability that a person will move out-of-state, out-of-region, or to a Pacific state using estimates from the discrete choice model described in Section 5. Standard errors are clustered at the ship-level.

Appendix – For Online Publication

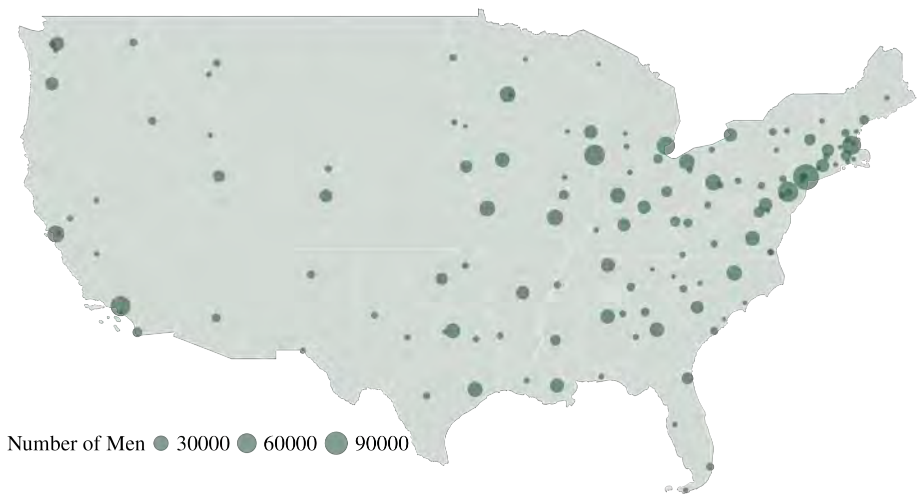
A Additional Figures

Figure A.1: Probability of living in different state than birthplace for white men



NOTES: This figure displays interstate geographic mobility for white men born 1890-1979. The dark green line shows the share of men living outside their state of birth for men ages 30-39, by birth decade. The lighter green line shows the share of men living outside their state of birth for men ages 10-19, by birth decade. The difference between these two lines is shown in Figure 1. Data source: Decennial census public use samples (1900-2010) and American Community Survey (2005-2024).

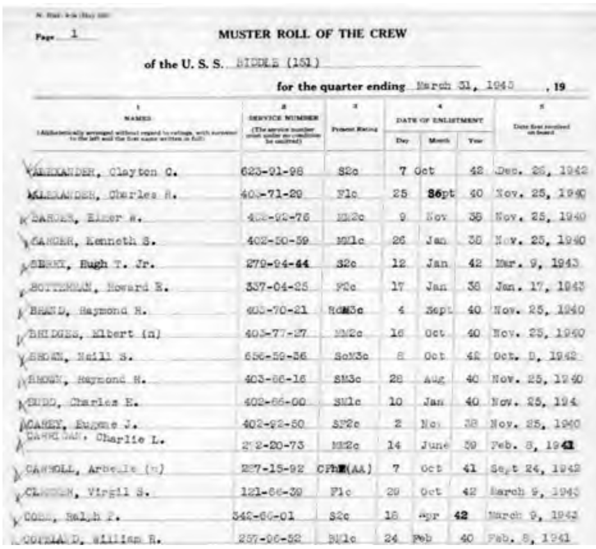
Figure A.2: Map of Cities of Enlistment

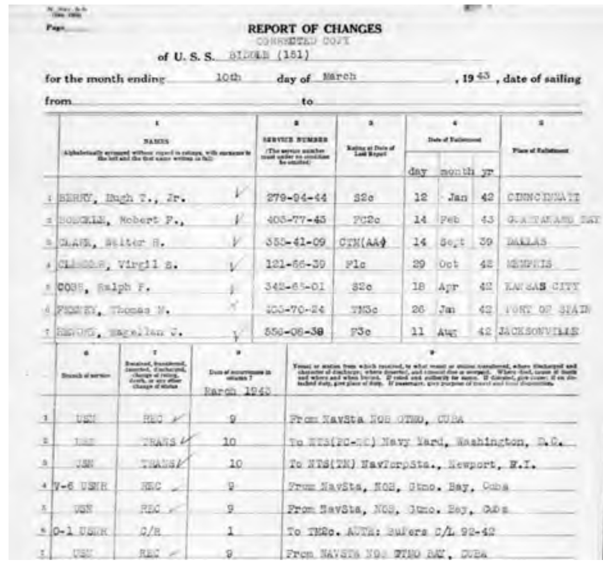


NOTES: This figure depicts the 133 cities where at least 100 sailors enlisted in the muster rolls. The size of each circle is scaled to the number of sailors who reported enlisting into the city. Data source: Place of enlistment is recorded either through direct report on the report of changes or through service number. Ranges of service numbers were assigned to each receiving station in the US. See B.2 for more details on data preparation.

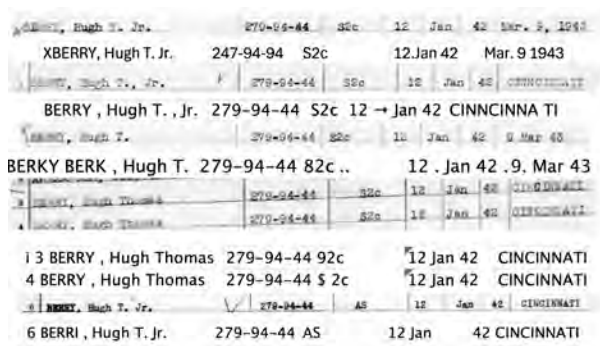
Figure A.3: Navy Muster Rolls

Panel A: Examples of Muster Roll and Quarterly Report of Changes for USS Biddle





Panel B: Example of Identifying "Hugh Berry" from Muster Rolls



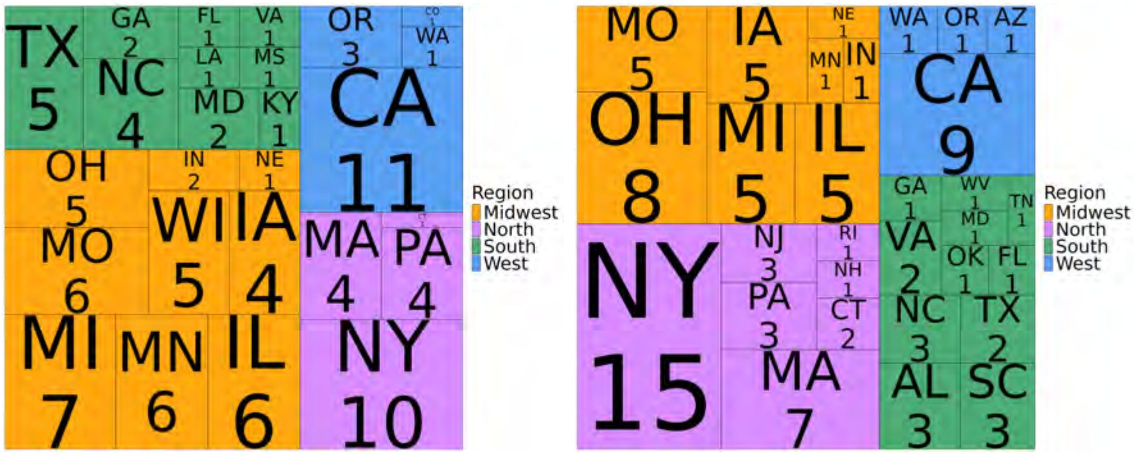
servno	lname	fname	mname	# scans	place	date e	date b
2799444	BERRY	HUGH	THOMAS	8	CINCINNATI OH	1-12-1942	3-9-1943

Panel C: Summary of Cleaned Muster Data

Metric	Value
Total number of people	1,456,484
Mean number of scans per person	9.6
Share on 1 boat	65%
Non-missing name	99%
Non-missing date enlist	61%
Non-missing place	96%
Non-missing ethnicity	93%
Non-missing HOH Occ. Score	89%

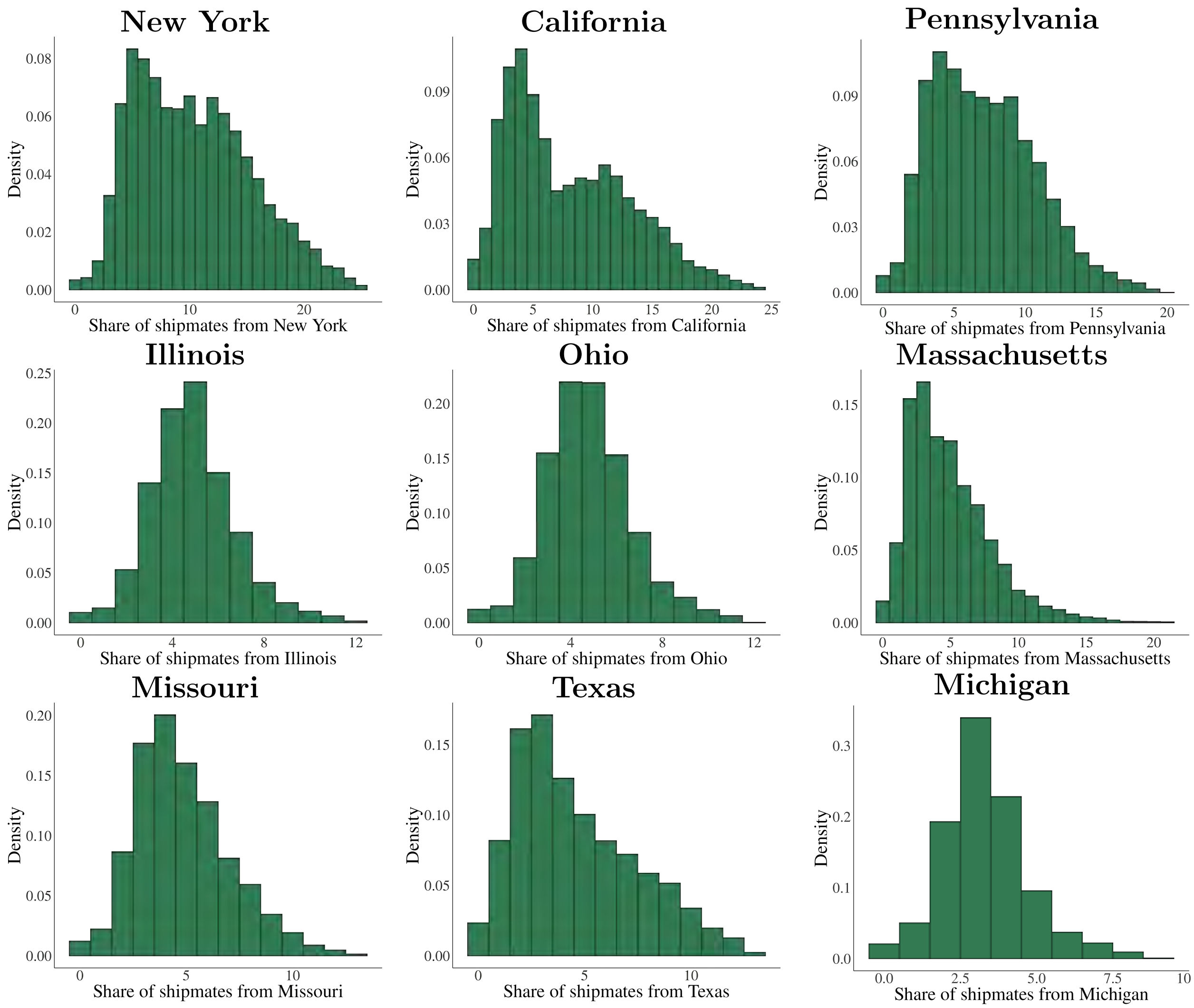
NOTES: Panel A displays example scans from muster rolls from the USS Biddle (DE-151). The left image is a quarterly census for for the quarter ending March 31, 1943, while the right image is a monthly report of changes from March 10, 1943. Panel B tracks all scanned entries and OCR transcriptions of Navy sailor Hugh Berry in the muster rolls, with the final row showing his cleaned characteristics post-processing. Panel C presents summary statistics for the cleaned data.

Figure A.4: Example Ship Exposure: USS Mero and USS Seacat



NOTES: This plot shows the geographic mix of shipmates for two different individuals on two different ships: the submarine USS Mero and the submarine USS Seacat. For each destination state, the number of distinct individuals each person is exposed to is represented on the plot.

Figure A.5: Shipmate Distribution by State



NOTES: This figure displays histograms of shipmate composition for the nine states with the largest representation in the Navy muster roll data. Each panel shows the distribution across individuals of the share of their shipmates who originated from that state. For example, the New York panel shows what fraction of each sailor's shipmates were from New York.

Table A.1: Summary Table of Ship Characteristics

	All Ships	Minesweepers	Destroyers
Number of Ships	5,019	618	381
Median Num. Ppl/Ship/Quarter	84	46	345
Median Num Quarters/Ship	7	9	10
Characteristics of Shipmates: Median [IQR]			
Num. of States	26.6 [15.50]	20.9 [10.0]	37.5 [5.9]
Share West (%)	11 [15]	12 [19]	11 [16]
Share North (%)	28 [20]	28 [23]	30 [18]
Share Midwest (%)	27 [10]	28 [23]	26 [18]
Share South (%)	27 [10]	26 [9]	27 [9]
# Ethnic Groups	15.6 [5.60]	13.1 [4.1]	18.0 [0.3]
Mean Occ Score	25.6 [1.03]	25.5 [1.2]	25.5 [0.6]
Occ in Farming (%)	18 [5]	19 [6]	19 [4]

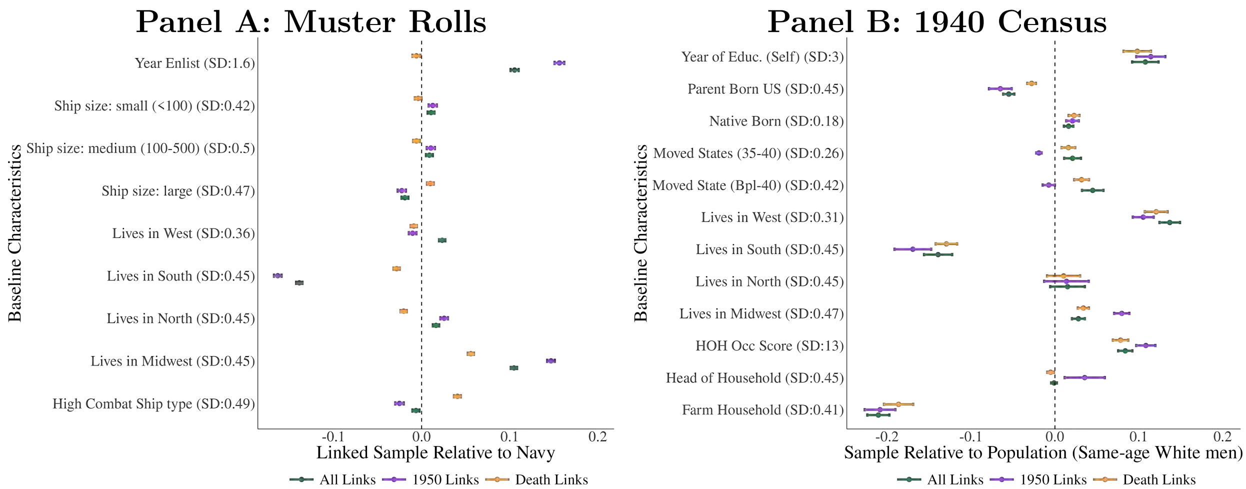
NOTES: This table presents summary statistics for Navy ships active between 1942 and 1945. Column 1 reports statistics for all ships in the sample, while columns 2 and 3 focus on two specific ship categories: minesweepers and destroyers. The first panel reports ship-level statistics. The second panel presents the distribution of shipmate characteristics, reporting the median and interquartile range [in brackets] across ships. Geographic shares reflect the distribution of sailors' pre-war state of residence across four Census regions. Ethnic groups are identified using name-based ethnicity classification following Abramitzky, Boustan, and Eriksson (2020). Occupational score and share in farming are constructed using characteristics of heads of household with the same name and state of residence in the 1940 Census. Ships with fewer than 10 sailors in a quarter are excluded from the sample.

Table A.2: Summary of Linked Sample

	Navy	Census 1940	Numident	FindAGrave
Pre-link restricted sample	1,355,514	26,465,390	14,734,046	1,836,012
Bilateral Links:				
Census 1940	343,330	–	–	–
Numident	428,213	4,420,699	–	–
FindAGrave	202,436	473,798	–	–
Number of Links from Muster Rolls:				
	1940 Census	1950 Census	Death (Any)	
Number of Links	478,000	266,000	578,000	

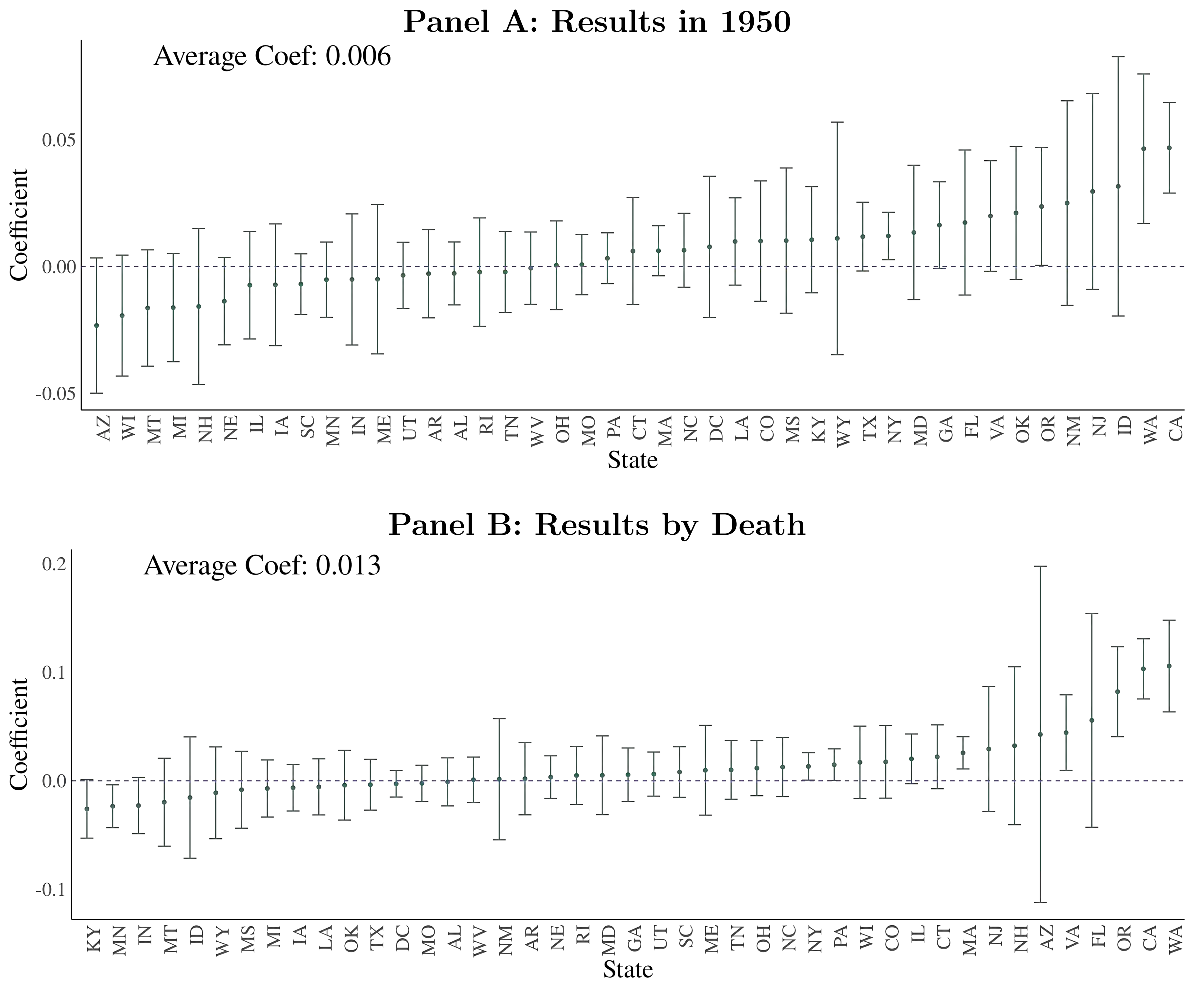
NOTES: Panel A compares three linked samples to the general population of rank-and-file personnel aboard Navy ships. Comparisons are made for three separate samples: All Links (Navy men linked to 1940 Census), 1950 Links (Navy men linked to 1940 and 1950 Census), and Death Links (Navy men linked to 1940 and death). Each row reports the estimates from a separate estimation; for each variable the characteristic is regressed on being in the linked sample. The coefficient for this is reported scaled to a standard deviation of the outcome variable. Panel B replicates this procedure but comparing linked individuals to same-age white men in the 1940 Census.

Figure A.6: Benchmarking Linked Sample



Notes: This figure compares three linked samples to the general population of rank-and-file personnel aboard Navy ships. Panel A compares characteristics between the linked sample and all men in the muster rolls data. Panel B compares characteristics between the linked sample and white men aged 16-40 in the 1940 Census. All Links refers to Navy men linked to the 1940 Census, 1950 Links refers to Navy men linked to both 1940 and 1950 Census records, and Death Links refers to Navy men linked to both 1940 Census and death records. Each row reports the estimates from a separate regression where the characteristic is regressed on being in the linked sample. The coefficient is reported scaled to a standard deviation of the outcome variable.

Figure A.7: Directed Migration to States



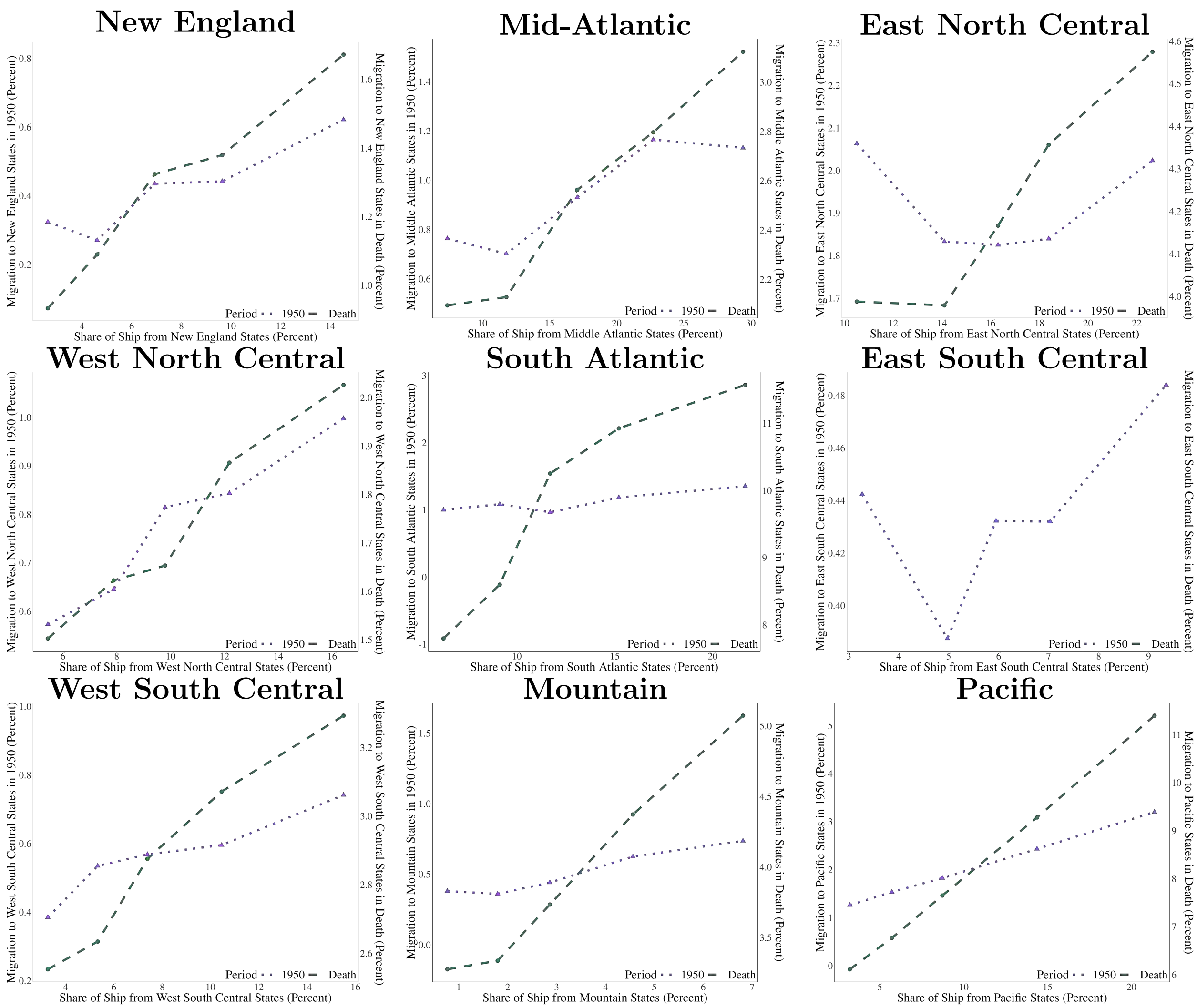
Notes: These plots show results from Equation (3) at the state-level. For each state depicted on the x-axis, the coefficient represents regressing the likelihood someone moves to state on the share of shipmates from that state, excluding individuals originally from that state. Along with the point-estimate, the 95% confidence interval is reported for each state. Standard errors are clustered at the ship-level. Panel A reports estimates for all 48 contiguous state in 1950, while Panel B reports estimates by death. The average coefficient is reported at the top of each plot.

Figure A.8: Directed Migration to Census Regions: Coefficients

	1950				Death			
	Midwest	North	South	West	Midwest	North	South	West
Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Share of Ship from Region	0.002 (0.007)	0.012*** (0.003)	0.018** (0.008)	0.010*** (0.003)	0.010 (0.011)	0.026*** (0.005)	0.029* (0.015)	0.046*** (0.005)
Sample	Non-Midwest	Non-North	Non-South	Non-West	Non-Midwest	Non-North	Non-South	Non-West
Dep Var Mean	0.022	0.012	0.018	0.007	0.053	0.033	0.140	0.040
Observations	117,493	129,518	142,318	169,242	172,089	181,944	188,154	233,820
R^2	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.03
Within R^2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
State FE	✓	✓	✓	✓	✓	✓	✓	✓
Category FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓

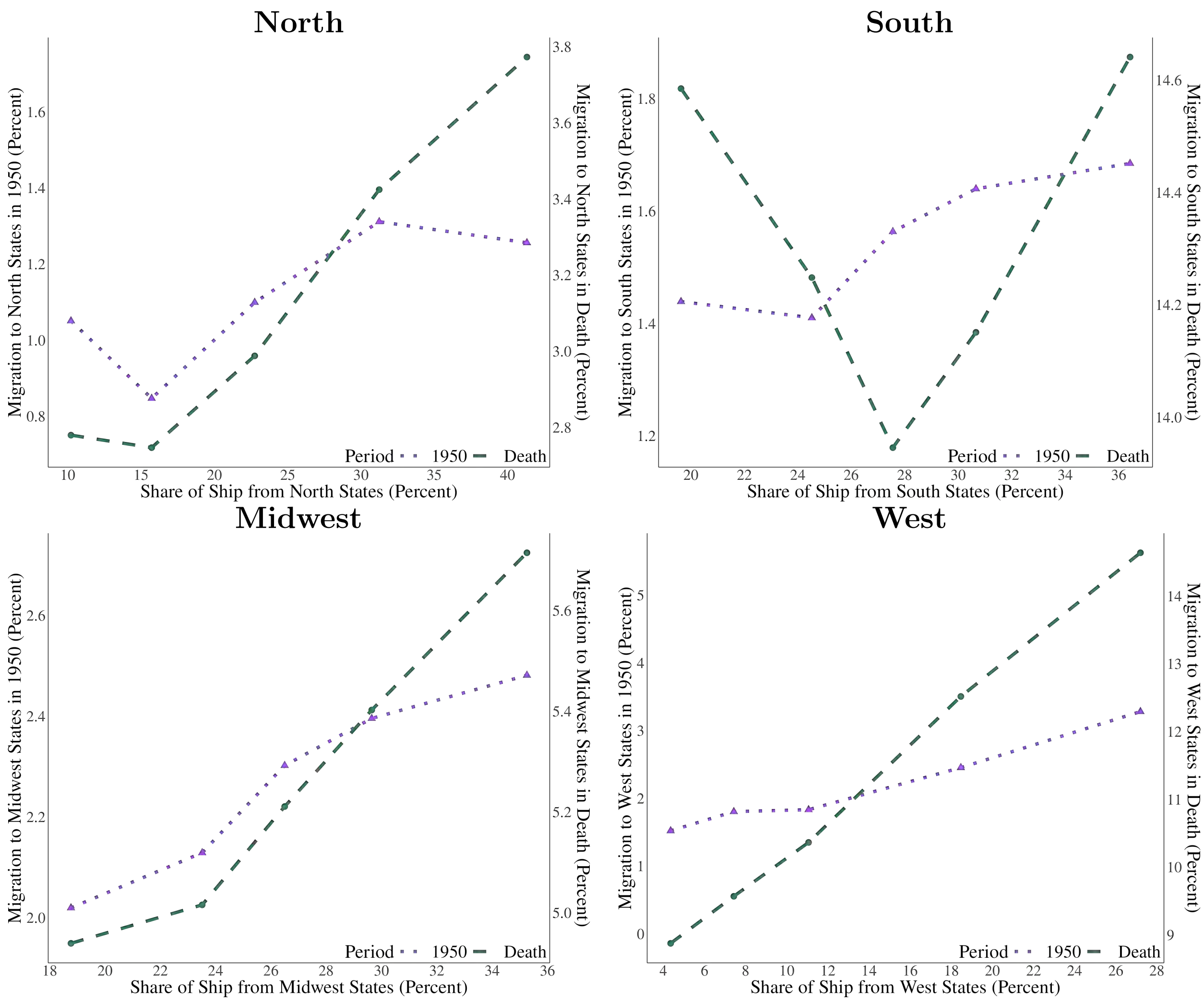
NOTES: This plot is analogous to Figure 4 except it shows directed migration at the regional level. This figure shows the impact of Navy ship networks on directed migration across Census region, by showing estimates from equation (3). The table presents $\beta^{j,t}$ showing the effect of increased exposure to shipmates from each Census region on migration to that region, measured both in 1950 and at time of death. Standard errors are clustered at the ship-level.

Figure A.9: Directed Migration (Census Divisions)



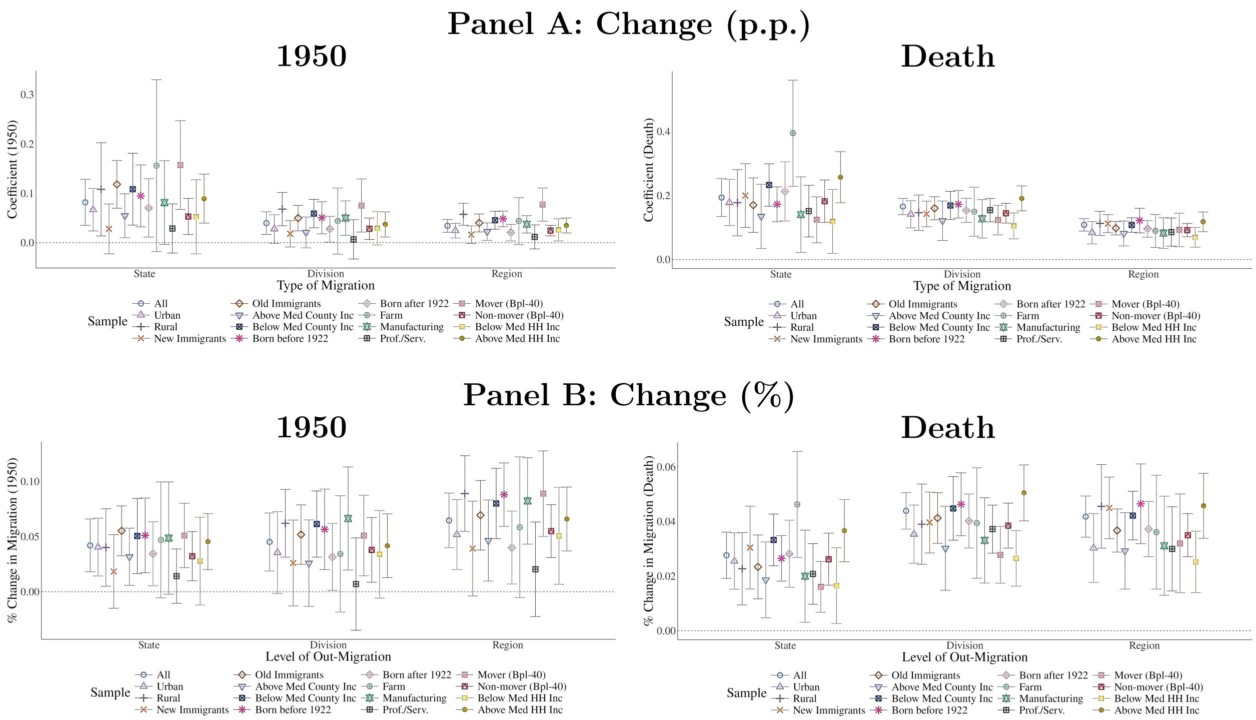
NOTES: This plot is a continuation of 4. Each plot shows the relationship between unconditional migration to Census divisions and shipmate characteristics. Each Census divisions is represented in a separate plot displaying unconditional migration by 1950 (purple, left axis) and by death year (green, right axis). Both axes span an equal range but are re-leveled to accommodate different baseline migration rates in each period. The average unconditional migration rate is calculated for each quintile for each home state, with a sample population-weighted average across home states reported on the plot.

Figure A.10: Directed Migration to Census Regions



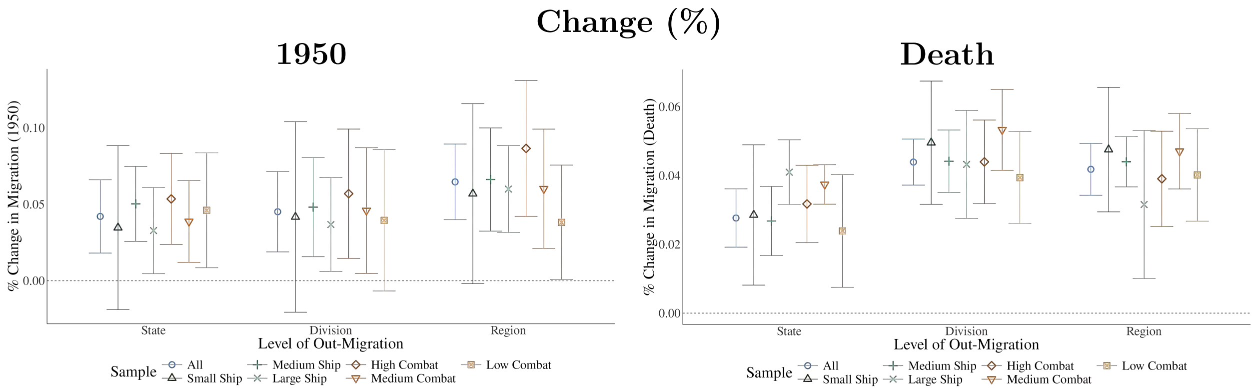
Notes: This plot is a continuation of A.8. This plot is analogous to Figure A.9 except it shows directed migration at the regional level. Each plot shows the relationship between unconditional migration to Census regions and shipmate characteristics. Each Census region is represented in a separate plot displaying unconditional migration by 1950 (purple, left axis) and by death year (green, right axis). Both axes span an equal range but are re-leveled to accommodate different baseline migration rates in each period. The average unconditional migration rate is calculated for each quintile for each home state, with a sample population-weighted average across home states reported on the plot.

Figure A.11: Heterogeneity in Out-migration by Baseline Characteristics



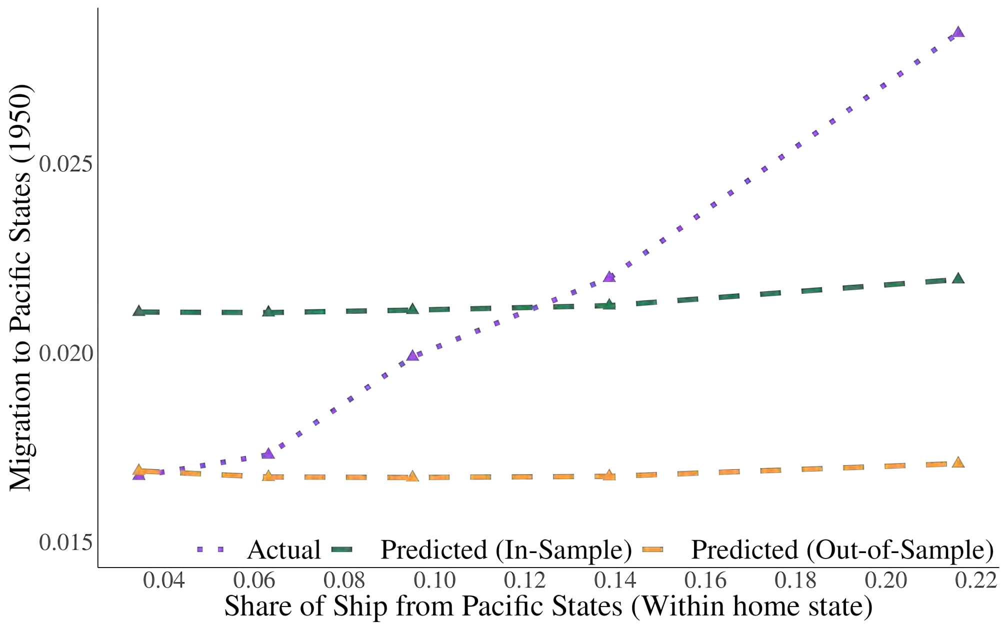
NOTES: This plot depicts results from Equation (2) for different sub-samples of the population. Each coefficient depicted shows the result of a single regression restricted to a specific sub-population. On each plot, coefficients are reported for 16 different samples and three different levels of geography. Panel A, reports the raw coefficient, and in Panel B the coefficient is scaled to report percent change from a 1 SD increase in exposure. The 95% confidence interval is also shown for each coefficient. Standard errors are clustered at the ship level.

Figure A.12: Heterogeneity in Out-migration by Ship Type



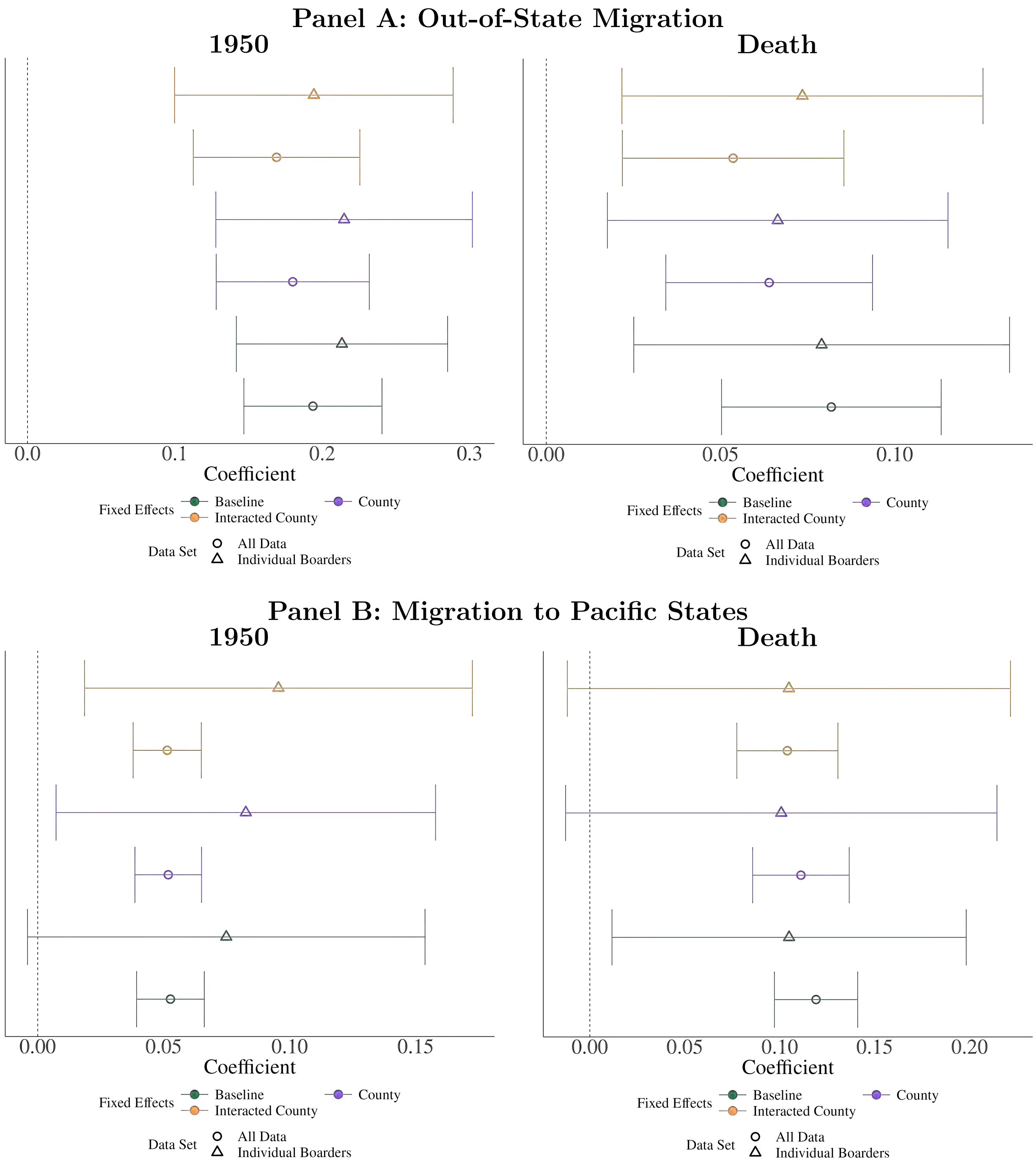
NOTES: This plot depicts results from Equation 2 for different sub-samples of the population. Each coefficient depicted shows the result of a single regression restricted to a specific sub-population. On each plot, coefficients are reported for 16 different samples and three different levels of geography. Panel A, reports the raw coefficient, and in Panel B the coefficient is scaled to report percent change from a 1 SD increase in exposure. The 95% confidence interval is also shown for each coefficient. Standard errors are clustered at the ship level.

Figure A.13: Robustness: Predicted vs. Actual Migration to Pacific States



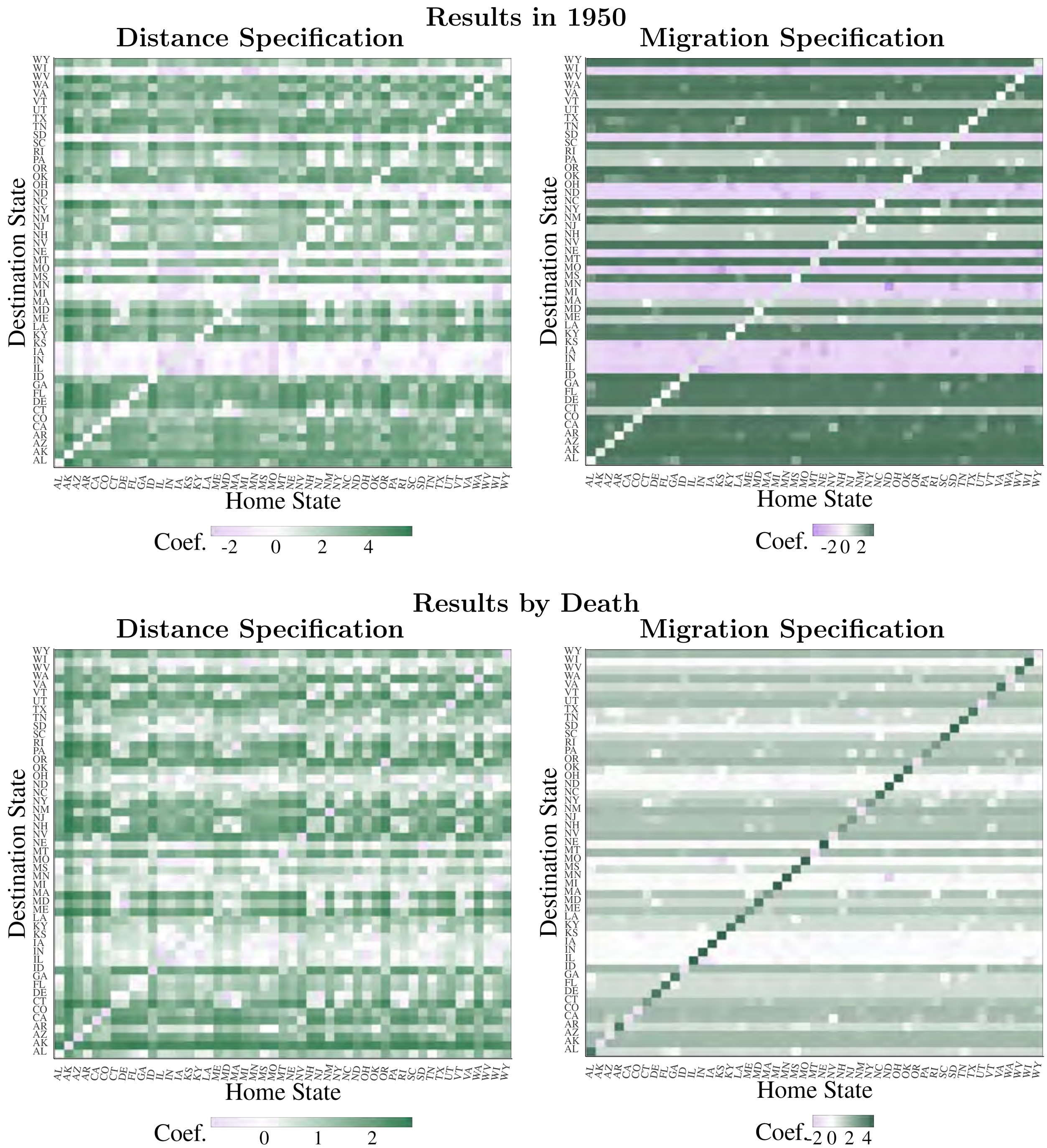
NOTES: This figure demonstrates the relationship between predicted and actual migration to Pacific states in 1950. Individuals from the same home state h in the linked Navy sample are are separated into five equal sized bins by the share of their shipmates from Pacific states. For each person, migration to Pacific states is predicted using pre-war characteristics from the 1940 Census (age, education, household size, home ownership, farm status, occupational score, and state and county of residence). This predicted migration is generated from two samples: individuals in the Navy sample (in-sample,green) and white men born between 1905 and 1928 (out-of-sample,yellow). The purple dotted line plots actual 1950 migration to Pacific states against these predicted values and is constructed analogously to plots in Figure A.9. For each of the three series–predicted (in-sample,green), predicted (out-of-sample,yellow), and actual (purple)–the average unconditional migration rate is calculated for each quintile for each home state, with a sample population-weighted average across home states reported on the plot.

Figure A.14: Robustness Exercises



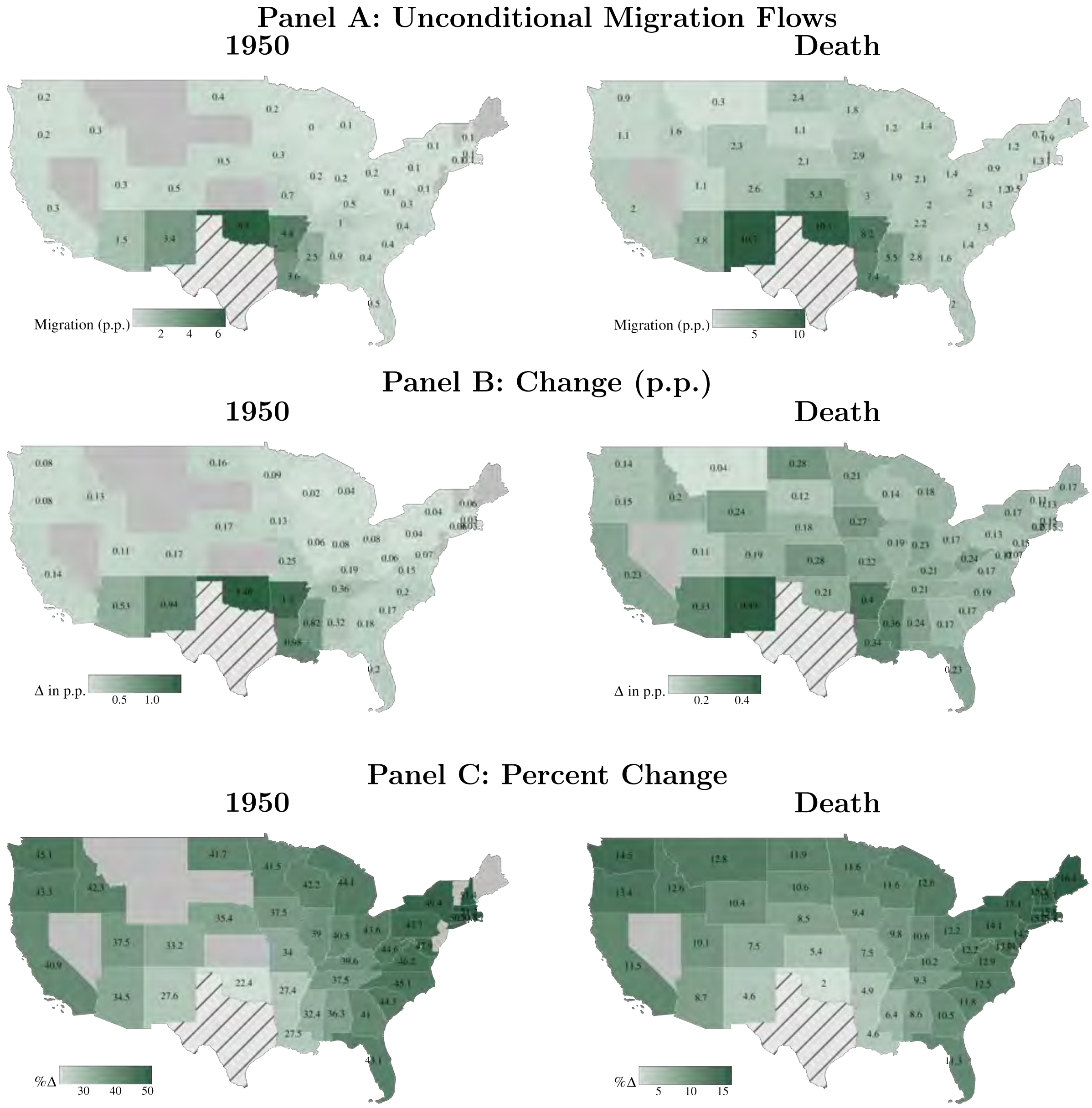
Notes: This figure presents robustness checks for the main migration estimates. Panel A shows estimates from Equation (2) for out-of-state migration, while Panel B shows estimates from Equation (3) migration to Pacific states. Each panel reports estimates for both 1950 outcomes (left) and outcomes at time of death (right). Each point represents a separate regression estimate under different specifications and sample restrictions. The baseline specification includes pre-war state, quarter-of-enlistment, and ship type fixed effects. Additional specifications add: (i) county-level fixed effects and (ii) fully interacted fixed effects (with county). Sample restrictions include: (i) limiting to individuals whose exposure to same-day boarders is less than 10%. Horizontal bars represent 95% confidence intervals with standard errors clustered at the ship level.

Figure A.15: Heatmap of Gravity Coefficients β_{hd}^{dest} and β_{d}^{home}



Notes: This figure displays heatmaps of the gravity model coefficients estimated from equation (4). Each cell represents a home-destination state pair, with color intensity indicating the strength of network effects. The top row shows coefficients for 1950 outcomes, while the bottom row shows coefficients for outcomes at time of death. The left panels show coefficients from the distance specification, where network effects are parameterized as a function of log distance between states. The right panels show coefficients from the migration specification, where network effects are parameterized using pre-existing migration flows between 1935 and 1940. Diagonal elements represent β_{d}^{home}, measuring the impact of ties to one's home state. Off-diagonal elements represent β_{hd}^{dest}.

Figure A.16: Impact of a 10 p.p. Increase in Texas Share on Migration

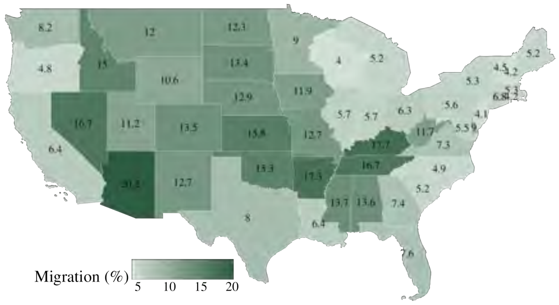


NOTES: This figure is an analog to Figure 6, but for measuring the impact of increased exposure to Texan shipmates on migration to Texas from each origin state. The counterfactual compares the predicted migration response from serving on a high-exposure ship (10 percentage points higher) to a low-exposure ship as described in Equation (8) using discrete choice estimates reported in Figure 5. Panel A reports average unconditional migration flows to California by origin state by 1950 and by a person's death. Panel B shows the increase in migration probability of serving on a high-exposure ship relative to a low-exposure ship in percentage points. Panel C displays the percent increase in migration probability relative to the average unconditional migration rate (Panel B/Panel A). The left map in each panel reports results in 1950, while the right map in each panel reports results by death.

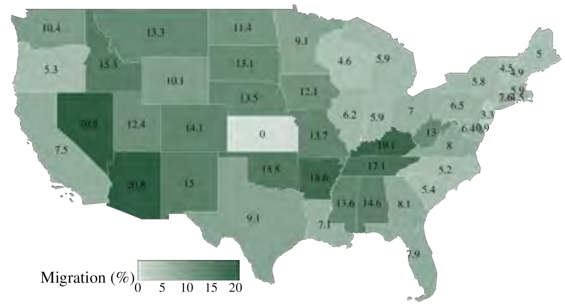
Figure A.17: Impact of 1 SD increase in ship exposure on division migration, by state of origin

Panel A: Average Out-Migration (p.p.)

1950

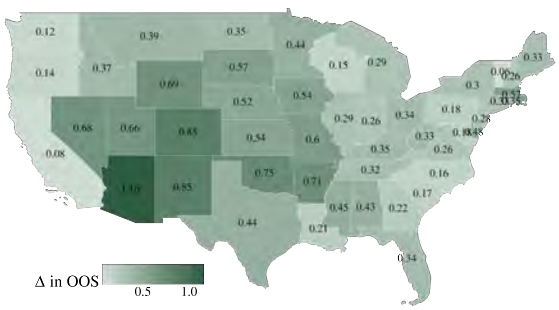


Death

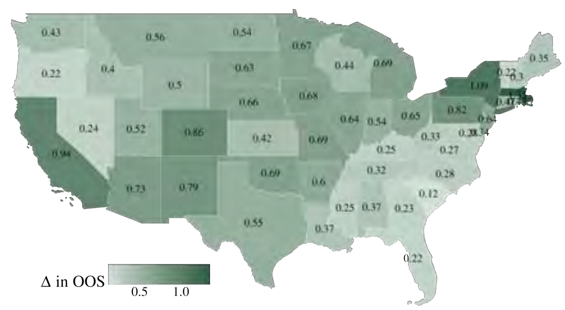


Panel B: Change from 1 SD Increase (p.p.)

1950

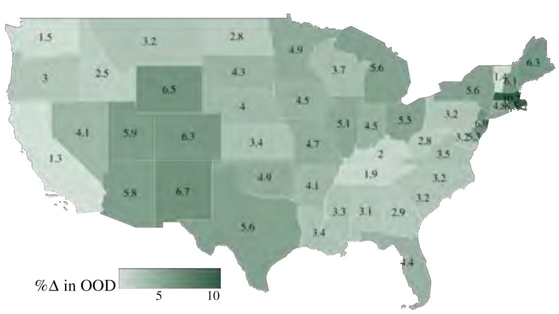


Death

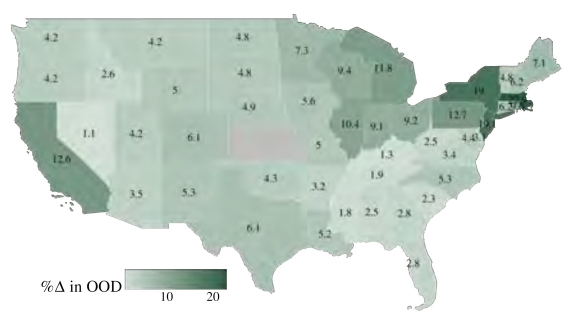


Panel C: Percent Change from 1 SD Increase

1950

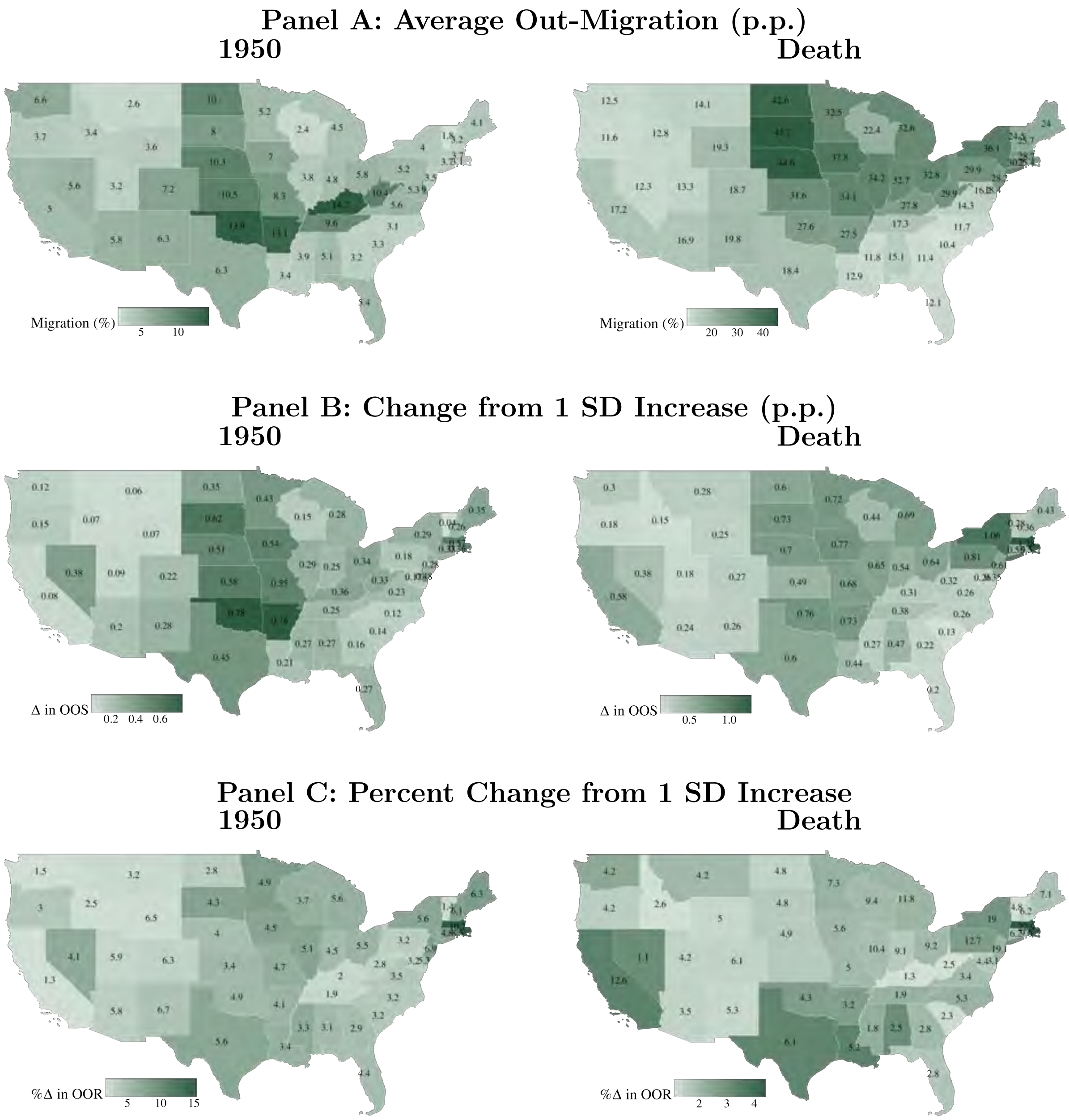


Death



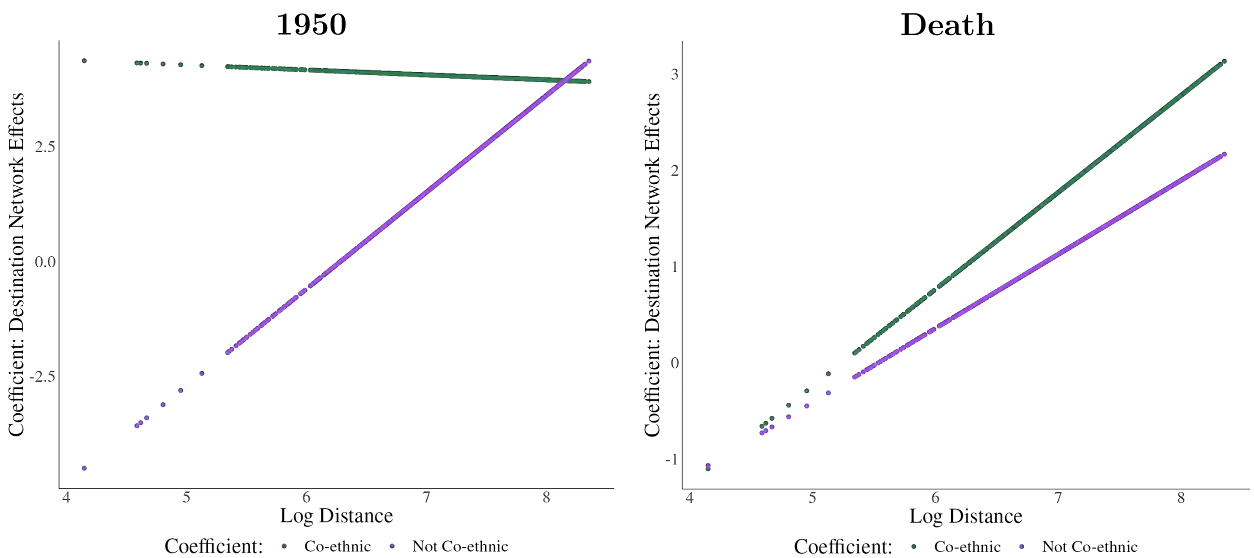
NOTES: This figure is analogous to Figure 7, but shows results for migration across Census divisions. Following Equation (9) and using estimates from Figure 5, the counterfactual, $(\Delta P_{div_d \neq div_{h(i)}|it})$, computes the difference in probability of moving out-of-division when assigned to one's actual ship versus a ship with average composition for their pre-war residence state $h(i)$. Panel A displays average out-of-division migration rates in 1950 and by time of death for men in the linked sample. Panel B reports one standard deviation in $(\Delta P_{div_d \neq div_{h(i)}|it})$ across individuals from state h, representing the change in migration probability from a one standard deviation increase in effective shipmate variation. Panel C shows this effect as a percentage change relative to average out-migration rates from each state (Panel B/Panel A). Each panel presents results for both 1950 (left) and time of death (right).

Figure A.18: Impact of 1 SD increase in ship exposure on region migration, by state of origin



NOTES: This figure is analogous to Figure 7, but shows results for migration across Census regions. Following Equation (9) and using estimates from Figure 5, the counterfactual, $(\Delta P_{reg_d \neq reg_{h(i)}|it})$, computes the difference in probability of moving out-of-region when assigned to one's actual ship versus a ship with average composition for their pre-war residence state $h(i)$. Panel A displays average out-of-region migration rates in 1950 and by time of death for men in the linked sample. Panel B reports one standard deviation in $(\Delta P_{reg_d \neq reg_{h(i)}|it})$ across individuals from state h, representing the change in migration probability from a one standard deviation increase in effective shipmate variation. Panel C shows this effect as a percentage change relative to average out-migration rates from each state (Panel B/Panel A). Each panel presents results for both 1950 (left) and time of death (right).

Figure A.19: Co-ethnics and Network Formation



NOTES: This figure plots the predicted coefficient β_{hd}^{dest} from equation (11) against logged distance between states, separately for co-ethnic and different-ethnicity shipmates. The left panel shows results for migration by 1950, while the right panel shows results for migration by time of death. Logged distance between states varies from 4.1 (RI to MA) to 8.4 (ME to CA), with the median origin-destination pair having logged distance of 7.3 (MA to WI).

Table A.3: Network Formation: Occupational Proxmity and Pre-War Occupational Score

	Rating Group		Occupation Score	
	1950	Death	1950	Death
β_{dest}	-8.39***	-4.42***	-9.75***	-4.45***
	(2.18)	(1.14)	(1.96)	(1.04)
$\beta_{logdist}$	1.53***	0.84***	1.70***	0.83***
	(0.31)	(0.16)	(0.28)	(0.14)
β_{rg}	0.21**	0.29***		
	(0.23)	(0.11)		
$\beta_{rg\ logdist}$	-0.03	-0.05**		
	(0.05)	(0.02)		
β_{occ}			-0.00	-0.00
			(0.02)	(0.01)
$\beta_{occ\ logdist}$			0.00	0.00
			(0.00)	(0.00)
β_{home}	-0.14**	0.12***	0.13	0.34***
	(0.25)	(0.12)	(0.22)	(0.10)
$\beta_{home\ rg}$	0.21***	0.21***		
	(0.04)	(0.02)		
$\beta_{home\ occ}$			-0.01	0.00
			(0.00)	(0.00)
R^2	0.65	0.41	0.65	0.41
N	4,348,954	6,756,488	5,536,734	8,430,555

Notes: This table reports discrete choice estimates from Equation (C.1) (Columns 1 and 2) and Equation (C.2) (Columns 3 and 4). Columns 1 and 3 report coefficient estimates for 1950, while columns 2 and 4 report esimates by time of death. Coefficients are estimated by Poisson-Pseudo Maximum Likelihood (PPML) estimation. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure A.20: Returns to Migration IV: Balance

	Migration Instrument (in SD)					
	State (1)	1950 Region (2)	Pacific (3)	State (4)	Death Region (5)	Pacific (6)
Native Born	0.001	-0.01*	-0.0007	-0.001	-0.004	0.001
	(0.007)	(0.007)	(0.006)	(0.008)	(0.005)	(0.005)
HOH Education	0.0003	0.0002	0.0002	-0.0004	-7.9×10^{-5}	-0.0004
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0003)	(0.0003)
HOH Occ Score	1.8×10^{-5}	0.0001	-8.7×10^{-5}	0.0001	2.6×10^{-5}	-0.0002*
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(8.5×10^{-5})	(8.5×10^{-5})
Age	0.001	0.001	9.8×10^{-5}	-8.5×10^{-5}	-3.7×10^{-5}	-0.0006
	(0.002)	(0.002)	(0.001)	(0.001)	(0.0009)	(0.0009)
Age Squared	-2.2×10^{-5}	-2.4×10^{-5}	2.4×10^{-6}	-9.8×10^{-7}	-2.3×10^{-6}	1.6×10^{-5}
	(3.3×10^{-5})	(3.4×10^{-5})	(2.9×10^{-5})	(2.8×10^{-5})	(2×10^{-5})	(2×10^{-5})
Moved States (35-40)	0.008	0.01**	0.01	0.01*	0.01***	0.007*
	(0.006)	(0.006)	(0.007)	(0.005)	(0.004)	(0.004)
Observations	140,278	140,277	124,641	140,877	140,877	124,048
R^2	0.80850	0.83229	0.87108	0.85534	0.92943	0.94543
Within R^2	2.72×10^{-5}	7.45×10^{-5}	5.89×10^{-5}	4.82×10^{-5}	9.96×10^{-5}	0.00011
1940 State by County FE	✓	✓	✓	✓	✓	✓
Ship type FE	✓	✓	✓	✓	✓	✓
First Quarter FE	✓	✓	✓	✓	✓	✓

NOTES: This tables show the underlying variation between the migration instruments defined in Section 7 and individual baseline characteristics in 1940. The dependent variable is the migration instrument normalized into standard deviations. The instrument is constructed as the predicted probability that a person will move out-of-state, out-of-region, or to a Pacific state using estimates from the discrete choice model described in Section 5. Columns 1-3 show results using migration instruments constructed for 1950 outcomes, while columns 4-6 show results using instruments for lifetime migration. The baseline characteristics are measured using the 1940 Census. Native Born is an indicator for U.S. birth, HOH variables refer to characteristics of the household head in 1940, and Moved States (35-40) indicates migration across state lines between 1935-1940. Standard errors are clustered at the ship-level.

Figure A.21: Returns to Migration IV: First Stage

	1950			Death		
	State	Region	Pacific	State	Region	Pacific
	(1)	(2)	(3)	(4)	(5)	(6)
Migration Instrument	0.76***	0.69***	0.85***	0.70***	0.72***	0.93***
	(0.04)	(0.06)	(0.09)	(0.04)	(0.05)	(0.09)
Observations	173,504	173,503	154,312	177,190	177,158	156,125
R^2	0.13156	0.13000	0.06327	0.05766	0.06743	0.05548
Within R^2	0.00290	0.00118	0.00199	0.00172	0.00166	0.00144
1940 State by County FE	✓	✓	✓	✓	✓	✓
Ship type FE	✓	✓	✓	✓	✓	✓
First Quarter FE	✓	✓	✓	✓	✓	✓

NOTES: This table shows the first stage from Equation (12). For each specification the migration instrument is constructed for that period t and migration type y, where instrument is constructed as the predicted probability that a person will move out-of-state, out-of-region, or to a Pacific state by time t using estimates from the discrete choice model described in Section 5. The dependent variable is an indicator for whether a person moved out-of-state, out-of-region, or to the Pacific by time period t. Columns 1-3 report estimates for migration in 1950, while columns 4-6 report estimates for lifetime migration. Columns 1 and 4 report estimates for migration out-of-state. Columns 2 and 5 report estimates for migration out-of-region. Columns 3 and 6 report estimates for Pacific migration where individuals original from the Pacific are dropped. Standard errors are clustered at the ship-level.

Figure A.22: Returns to Migration IV: Reduced Form

	1950 Occscore 1950 (logged)			Death Zipcode Inc Death (logged)		
	(1)	(2)	(3)	(4)	(5)	(6)
Migration Instrument	0.009**	0.03***	0.03***	0.31***	0.38***	0.65***
	(0.004)	(0.009)	(0.01)	(0.05)	(0.06)	(0.13)
Observations	173,504	173,503	154,312	144,203	144,202	127,270
R^2	0.94222	0.94222	0.94418	0.24403	0.24403	0.21607
Within R^2	3.11×10^{-5}	6.78×10^{-5}	3.82×10^{-5}	0.00038	0.00040	0.00026
State-county40 fixed effects	✓	✓	✓	✓	✓	✓
Category fixed effects	✓	✓	✓	✓	✓	✓
Time fixed effects	✓	✓	✓	✓	✓	✓

NOTES: This table shows reduced form estimates from from Equation (12). For each specification the migration instrument is constructed fof period t and migration type y, where instrument is the predicted probability a person will move out-of-state, out-of-region, or to a Pacific state by time t using estimates from the discrete choice model described in Section 5. The dependent variable is a proxy logged income in period t. In 1950, income is measured by occupational score of the individual. By time of death, income is measured by median household income in the zipcode a person resided in at death. Income at death is reported for the year 2000 in nominal dollars. Columns 1-3 report estimates for migration in 1950, while columns 4-6 report estimates for lifetime migration. Columns 1 and 4 report estimates for migration out-of-state. Columns 2 and 5 report estimates for migration out-of-region. Columns 3 and 6 report estimates for Pacific migration where individuals original from the Pacific are dropped. Standard errors are clustered at the ship-level.

Table A.4: Returns to Networked Migration (1950)

	Occscore 1950 (logged)					
		OLS			IV	
	(1)	(2)	(3)	(4)	(5)	(6)
State Mover	0.02***			0.01**		
	(0.0006)			(0.005)		
Region Mover		0.02***			0.04***	
		(0.0010)			(0.01)	
Pacific Mover			0.01***			0.03***
			(0.001)			(0.01)
Observations	175,275	175,275	155,995	173,504	173,503	154,312
F-stat				370.76	133.45	87.031
1940 State by County FE	✓	✓	✓	✓	✓	✓
Ship type FE	✓	✓	✓	✓	✓	✓
First Quarter FE	✓	✓	✓	✓	✓	✓

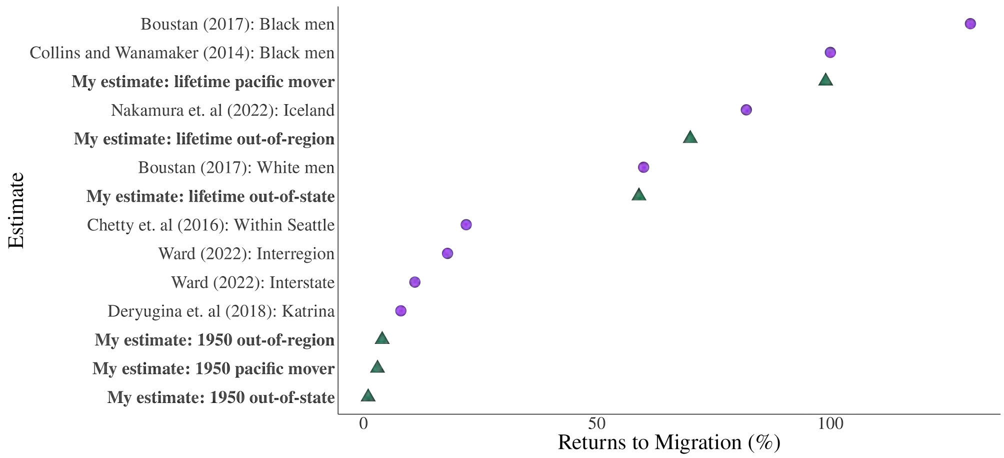
NOTES: This table reports coefficients from Equation (12) showing the returns to migration by time of death (lifetime). Columns 1-3 report OLS estimates, while Columns 4-6 report IV estimates. The outcome is logged occupational score of the person in the 1950 Census, where the occupational score as the average income for each occupation in 1950 in hundreds of 1950 dollars. Individuals without recorded occupations are or non-earning occupations are dropped from estimation. Income is reported for the year 2000 in nominal dollars. State mover is an indicator for if a person moved across state lines. Region mover is an indicator if a person moved between Census regions. Pacific mover is an indicator if someone who was not previously living in the Pacific Census division moved to a state in that division (excluding Alaska and Hawaii). In columns 4-6, the instrument is constructed as the predicted probability that a person will move out-of-state, out-of-region, or to a Pacific state using estimates from the discrete choice model described in Section 5. Standard errors are clustered at the ship-level.

Table A.5: Impact of Networked Migration on non-pecuniary outcomes

	HH Size (1950) (1)	Married (1950) (2)	Wife Different Birthplace (3)	Age at Death (4)
State Mover	-0.28	-0.04	0.29***	-0.45
	(0.41)	(0.06)	(0.07)	(1.2)
Observations	173,504	173,504	152,259	176,731
Mean of Dep Var	4.1	0.73	0.39	75
F-stat	370.76	370.76	340.79	296.59
State-county40 fixed effects	✓	✓	✓	✓
Category fixed effects	✓	✓	✓	✓
Time fixed effects	✓	✓	✓	✓

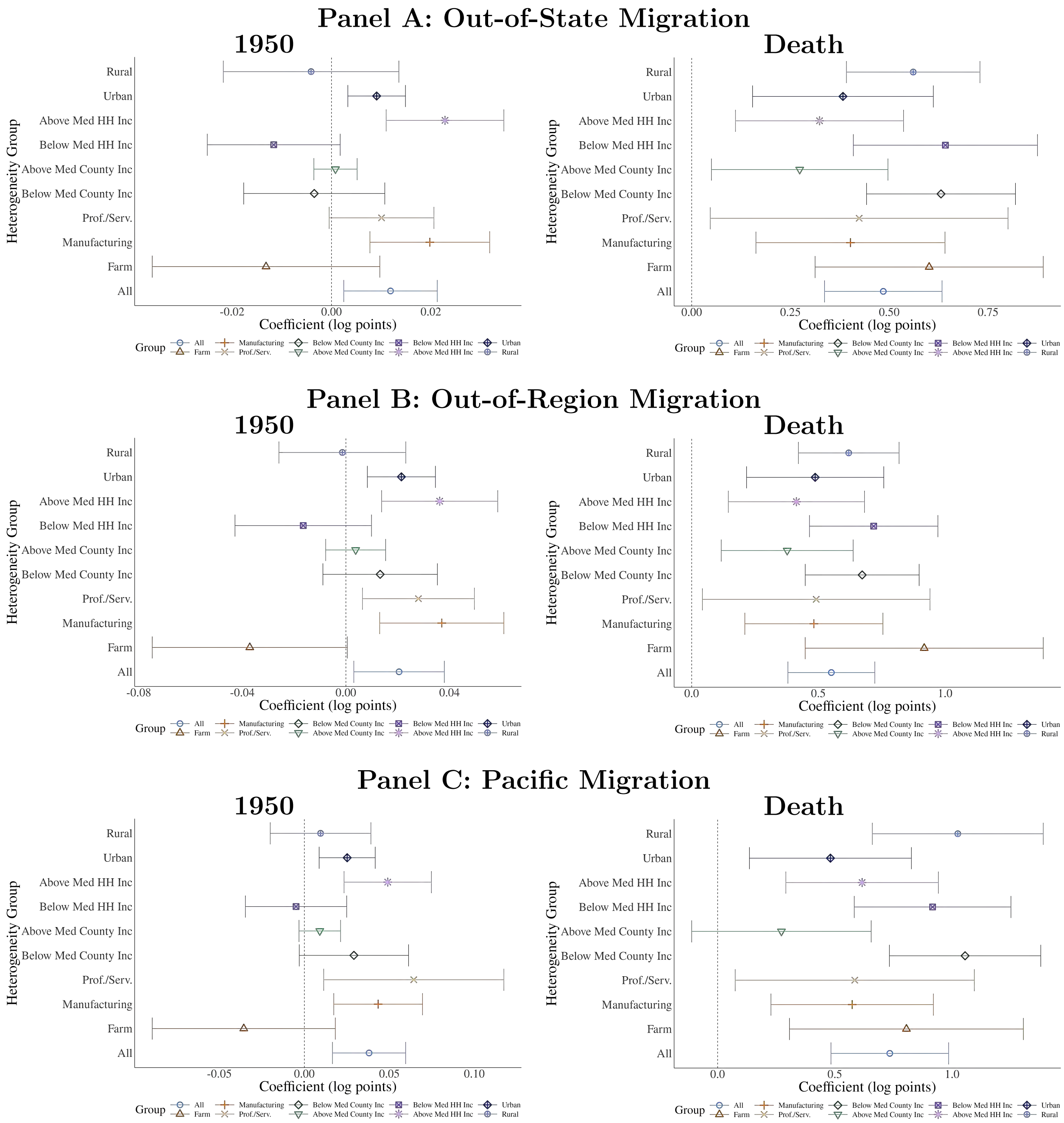
NOTES: This table reports IV estimates analogous to Equation (12) for non-pecuniary effects of networked migration. Each specification measures the impact of moving out-of-state, where the instrument is constructed as the predicted probability that a person will move out-of-state using estimates from the discrete choice model described in Section 5. Columns 1-3 report estimates for outcomes in 1950 and use the predicted probability of moving out-of-state by 1950, while column reports a lifetime outocme and uses the predicted probability of moving out-of-state by time of death. Column 1 reports the impact on total household size. Column 2 reports the impact for whether someone is married. Column 3 reports whether a married man's wife is born in either a different state or country. Finally, column 4 reports the impact of out-of-state migration on age of mortality. Standard errors are clustered at the ship-level.

Figure A.23: Benchmarking Returns to Migration against the Literature



NOTES: This figure presents estimates from Tables 1 and A.4, showing the corresponding percent change in income for people induced to migrate due to Navy networks. Additional estimates from various economics papers are included to provide comparisons of the pecuniary returns to migration. All calculations are the authors' own.

Figure A.24: Returns to Migration: Heterogeneity



NOTES: Each plot reports coefficients from Equation (12) for a given type of migration and a specific time period. On each plot, the specification is run restricting the population to different subsets, and for each subset the coefficient is report with a 95% confidence interval. Panel A reports estimates for returns to out-of-state migration. Panel B reports estimates for returns to out-of-region migration. Panel C reports estimates for returns to migration to Pacific states. All standard errors are clustered at the ship-level.

B Data Appendix

B.1 World War II Muster Rolls

The muster rolls data, available digitally from the National Archives catalog, contains 6,459,023 scans divided between 32,101 file units.[58] This data covers the near-universe of all "activities" within the Navy between January 1, 1939 and January 1, 1949. An "activity" is any unit within the Navy including ships, training centers, stations, etc. The muster rolls report lists of enlisted personnel formally attached to each ship or activity on a quarterly basis.

I download all scans and metadata available for each of the 32,000 file units.[59] Titles of file units typically identify the name of the ship or other activity as well as dates of coverage. File units can contain anywhere from 1 scan up to 2000 scans. Each scan can have metadata which varies from containing no information, to identifying the type of scan, to occasionally including volunteer transcriptions. Information from specific ships can be located across multiple file units. I restrict attention to the approximately 14,000 file units where the "activity" is identified as being a ship. Excluded activities include construction battalions, airborne units, administrative units, hospitals, and training centers.

B.1.1 Muster Rolls Cleaning Procedure

I use optical character recognition (OCR) to categorize and extract data from each scan. I use Google Vision OCR software and the Layout Parser python package. Within each file unit, I am interested in scans of quarterly muster rolls that identify all enlisted personnel attached to ship at the end of the quarter and monthly reports of changes that identify any personnel changes within a month. These two types of scans typically make up 40% of a file unit. Remaining scans include title pages, blank pages, and lists of passengers Using a combination of metadata and data contained within the headers of scans, I categorize scans as belonging to one of four categories: quarterly muster roll, monthly change reports, other types of scans, and scans I cannot categorize (henceforth identified as miscellaneous). I drop scans that I can confidently identify as not being a quarterly report of monthly change report. I finally extract characteristics from the tabular structure of each scan from quarterly muster rolls, monthly change reports, and miscellaneous scans. Once this process is completed, I have one csv of extracted table and header contents for each scan within a file unit.

I then clean these csvs to construct two ship-level datasets. First, for each ship, I extract the set of individuals identified by a unique seven-digit service number who were ever formally

[58]Record Group 24: Records of the Bureau of Navy Personnel. https://catalog.archives.gov/id/594996.
[59]Main access and download period between October 2022 and December 2022

attached to that unit. For each individual, I identify name, service number, where they enlisted, when the enlisted, and when they boarded the ship. Second, I construct a quarterly panel of each person attached to ship in each quarter as well as their Navy occupational rating code at the end of the quarter.

As quality scans are highly variable and OCR encoding errors are common, I use extensive cleaning processes to identify as high quality information about each individual as possible. The high-level approach to cleaning is twofold: first I correct for common encoding errors at the observation level. Table B.1 reports for each relevant field high-level cleaning strategies. Then I use multiple occurrences of individuals across scans to extract the highest confidence attribute for each individual. For instance, if a person with the name "Hugh T. Berry" appears on two different scans but there is a 1-digit difference in service number, I combine these two people and prioritize the version of the service number that is more common.

Table B.1: Encoding Errors and Cleaning Procedures

Field	**Common Encoding Errors**	**Cleaning Procedures**
Last name	Pen along left column leads to leading character (often "X"), extra characters, duplicated characters, misspellings of less common names	Use order within scan to identify when last names are out of alphabetical order, match to last names in the 1940 census to fix common encoding errors, remove leading characters, clean up common encoding errors due to compound last names.
First name	Misspellings, extra characters	Match to first names in 1940 census to fix common encoding errors, remove any unusual character types.
Middle name	Mixed representation: middle name vs. initials vs. suffixes, extra characters	Remove suffixes, separate initials from names, ignore observations with long character lengths suggesting encoding errors.
Service number	Number encoding error, part of service number appearing in other fields, duplicated numbers	Identify 7-digit patterns using common representations, search for alternative versions of the service number in other fields, remove non-numeric characters, remove instances of duplication
Rating	Extra characters, number and letter encoding errors	Use Navy documentation on rating codes to identify and correct common transcription errors for each rating type.
Date enlistment and boarded	Misspellings of month, unclear whether a number refers to a day or month, number encoding errors	Clean up years using a known valid date range, fix common transcription errors for months, and use context within the scan to distinguish whether ambiguous numbers refer to the month or the day.
Place of enlistment	Extraneous characters, misspellings. Ambiguous city names	For the 100 most common places of enlistment, identify transcription errors and apply fuzzy string matching to associate entries with the correct location.

During this process, I also incorporate information from tags imputed by volunteers in the catalog. When available the tags while list name, date of enlistment, and place of enlistment, but they do not report service number. I use fuzzy matching on names within scans to replace encodings with tag information when possible.

In the last part of the cleaning procedure I construct a quarterly panel. For each individual, I identify the first quarter they were on a ship from a mixture of the date they boarded and the first scan I see them. I identify the last quarter on board using the last scan I see them. The date attached to each scan is recovered in one of three ways: first it can be directly reported in the metadata, second it can be extracted from the OCR of the header (though often messy), or it can be extracted from the order scans existing within the file unit. To the last point, imagine that I know a ship within that file unit were active over the course of four quarters. If I know that this muster roll is placed such that it is the second quarter than I can back out the specific date belonging to that muster roll. For each person, I then interpolate between their first and last quarter aboard to construct a panel.

Once I have constructed cleaned versions of ship-level data, I proceed to bring the data together, and in particular, harmonize individual characteristics between ships. Within this process, I identify individual service numbers that appear on multiple ships and harmonize any discrepancies of characteristics about individuals that differ between ships. I also de-duplicate individuals who based off of name and service number similarity are the same but have been separated by small encoding errors. At this point, I drop any observations that are likely "false" – in particular, I drop service numbers that only appear one time in any scan.

Once completed, I have a final dataset that contains 1,450,000 people across 5,200 different ships.

B.2 Additional Data Sources

Place of Enlistments

Place of enlistment data was sourced from two primary methods: monthly reports of changes and service number blocks assigned to receiving stations. This dual approach allowed for comprehensive coverage and cross-validation of enlistment locations. Service number blocks were identified using several internal Navy documents:

- A Bureau of Personnel document (December 15, 1942) specifying blocks for 69 receiving stations.
- An update (August 27, 1943) providing additional ranges for stations that had exhausted their initial allocations.
- A 1990s Navy document detailing World War II service number ranges constructed from a broader set of records.

Some service number ranges (e.g., 1,000,000 to 2,000,000) could not be uniquely assigned to enlistment locations. These included numbers for World War I veterans who re-enlisted, though some were repurposed for new enlistments. The process for assigning place of enlistment was as follows:

1. Identify candidate locations using both report of changes and service number blocks.
2. When sources agreed but differed in geographic specificity, select the more granular location.
3. In cases of disagreement, prioritize the location indicated by service number blocks, as report of changes often reflected training locations rather than enlistment sites.
4. Use granular service number blocks (in 1,000-number increments) to infer additional assignments.

This methodology yielded place of enlistment data for over 95Figure A.2 illustrates the geographic distribution of enlistment locations that processed at least 100 sailors.

Dictionary of American Naval Fighting Ships

The *Dictionary of American Naval Fighting Ships* (DANFS) was a series of reference books containing facts about the service histories of Navy ships used by the US Navy. These reference books were written between 1959 and 1991 and published in nine volumes for every Naval vessel active over the history of the Navy's service. These volumes were digitized volunteers and made available online at https://www.hazegray.org/danfs/.

The following information is usually contained within en entry for each ship: name, specifications (size, complement, ship type, etc.) , dates of activity (commissioned, laid down, launched, major battle dates), information about places the ship went (where it was laid down, launched, areas or patrol and battles), battles and commendations (number of battle stars)

I scrape the record of every Naval vessel within the muster rolls data and use textual processing to extract the following pieces of information about each ship: complement, dates launched and sunk (if available), World War II battle stars, location laid down and launched, and theater of activity during World War II.

Navy Occupation Rating Data

Within the muster rolls data occupation and occupational rank is abbreviated by alpha-numeric codes. For instance, "S1C" corresponds to Seaman 1st Class. These ratings both provide information about the nature of the job an individual was doing on a given ship as well pay grade and hierarchy of that person relative to others. For instance "1st Class" ranks are one grade above "2nd Class" ranks within the same occupation.

I use documentation from various Navy sources to construct a crosswalk of abbreviations to occupational title, paygrade, and divisional branch.[60]

World War II Navy Casualty Records

I collect Navy casualty records from a few different sources. First, for each state and county, the National Archives contain scans of every individual man from that state and county who was either wounded, missing, or dead in combat.[61]

Second, I have additional information on Naval losses of particular vessels including the date and geographic location when the vessel was lost.[62]

Finally, I have records on the number of individual casualties (wounded, killed, missing) by category of ship type during the war.[63]

Full Count 1940 and 1950 Census

I use restricted versions of the 1940 Full Count Census and the public release of the 1950 Full Count Census. The 1940 Census was the first to include questions on income, supplementing existing data on occupation and employment status. It collected detailed residence information, including 1940 address and 1935 location (either the same address or a different location specified down to the town/city or county level).

The 1950 Census, while more limited in scope than its 1940 predecessor, asked 20 questions of all individuals, with additional questions in the sample-line section (20% of individuals). Reliability issues have been noted with both the veteran status question (Bailey, 2011) and with recorded income. For this reason neither variable will be used in subsequent analysis.

Numident Social Security Death File

I use the Numident Social Security Death File provided through the CenSoc project (Breen, Osborne, and Goldstein, 2023)/ The records are collected from the Social Security Death Master file that has been cleaned to the Berkeley Unified Numident Mortality Database (BUNMD). There are roughly 50 million records that cover a high coverage of all deaths in the United States between 1985 and 2007.

[60]Sources: https://www.cem.va.gov/docs/wcag/hmm/Inscription-Abbreviations-Ranks-Navy.pdf https://www.ibiblio.org/hyperwar/USN/ref/Ranks&Rates/index.html

[61]Source: https://www.archives.gov/research/military/ww2/navy-casualties/south-dakota.html

[62]https://www.navsource.org/Naval/losses.htm#ms

[63]https://apps.dtic.mil/sti/citations/ADA230803

FindAGrave

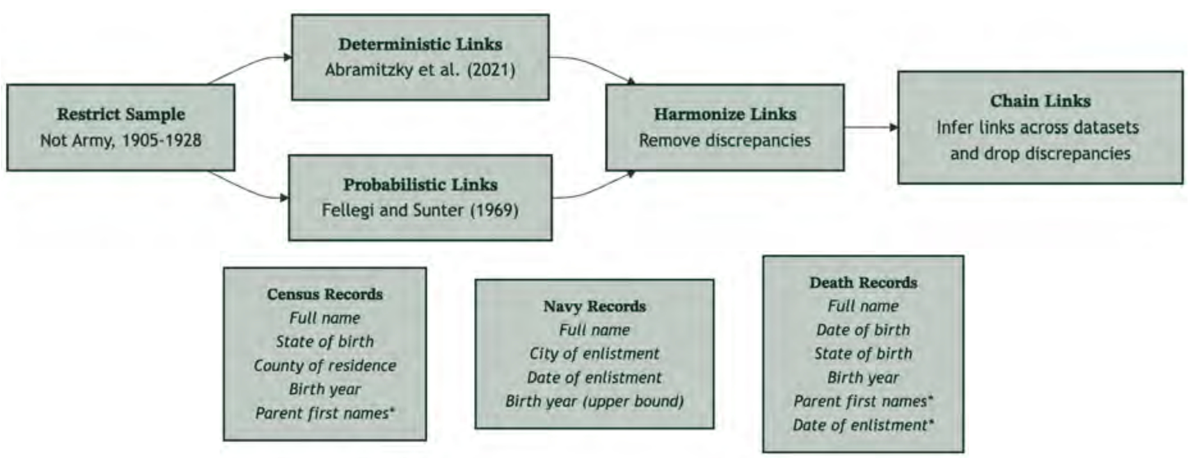
FindAGrave.com is the largest online database of gravestones. The website is primarily maintained by public contributions in which individuals will upload images of gravestones with transcriptions of information contained on those gravestones. For a portion of entries on the site, particular those belonging to veterans, information on gravestones are from large public databases maintained by military cemeteries such as Arlington or specific military agencies within the government.

I collect 2.8 million records belonging to individuals on FindAGrave.com as being born between 1900 and 1928, and additionally being tagged as a military veteran. Information almost always included on an individual include date of birth, date of death, cemetery location, and name. Very often, individual's gravestone pages will also include additional information on place of birth and place of death.

B.3 Additional Linking Information

This section elaborates on the four-step linking procedure discussed in Section 3. Figure B.1 provides a high-level visual overview of the linking procedure and the datasets connected in this process.

Figure B.1: Linking Procedure



Sample Restriction

To ensure accurate identification of unique matches, I first restrict the sample to individuals who could plausibly serve as rank-and-file personnel aboard a Navy ship. Across all datasets, I focus on men born between 1905 and 1928 who did not serve in the US Army. This date range encompasses the full spectrum of birth years for combat draft eligibility. According to enlistment records, over 99% of enlisted men in the Navy were born within this range. This restriction is particularly valuable for distinguishing between same-name father-son pairs.

To exclude men who served in the Army, I employ a conservative version of the linking procedure, connecting Army enlistment records to the 1940 Full Count Census, FindAGrave, and the Numident Social Security Death Records. The Army enlistment records contain additional high-quality linking fields such as state of birth and year of birth, which facilitate more accurate linking.

Following these restrictions, I proceed to construct bilateral links between each dataset.

Deterministic Linking

I implement a version of deterministic linking pioneered by Abramitzky et al. (2021). This approach identifies unique linked observations that match on several high-quality characteristics. Figure B.1 illustrates the linkable fields available in each dataset. For linking between Navy records and Census records, I rely solely on unique names. I employ a fuzzy name buffering approach, allowing for unique name matches within a Jaro-Winkler string distance of 0.05. To reduce false positives, I require that matches occur within a narrower fuzziness while remaining unique in a more general fuzzy band (e.g., 0.10). When linking Navy records to death records, I incorporate date and year of enlistment as additional potential linking fields, while maintaining the same fuzzy uniqueness procedure for names. For links between death records and census records, I add standard deterministic framework linking variables: state of birth and year of birth (allowing for some fuzziness).

Probabilistic Linking

I implement the Fellegi and Sunter (1969) method using the Python package `splink`. This framework estimates a model that predicts whether any two records belong to the same individual, given the set of characteristics they do and do not match on. This linking style is ideal when datasets have various linking variables that are either continuous or of variable quality, making the construction of a deterministic decision tree infeasible. I utilize all the linking fields from the deterministic approach, with some additions. For linking between Navy records and the 1940 Census, I include the physical distance between the county of residence in 1940 and the city of Navy enlistment. I also incorporate a probabilistic age component based on the year of enlistment. When linking Numident records to the Full Count Censuses, I add characteristics of the mother's and father's names. To address concerns about over-clustering, particularly on household and enlistment date, I remove links with large discrepancies in specific fields such as first name and last name after constructing initial candidate links. I then restrict to only unique links.

Harmonization and Chaining

After generating bilateral probabilistic and deterministic links between each dataset, I harmonize across linking procedures and perform a final round of link chaining to infer additional connections.

In the harmonization process, I remove any discrepancies where the links established between individuals disagree. Less than 0.3% of all generated links create such discrepancies. Both probabilistic and deterministic linking procedures provide informative unique links. Table B.2 shows the share of links between each bilateral linking dataset created from either procedure or both procedures.

Table B.2: Share of Link Between Datasets Generated by Each Procedure

	Probabilistic	**Deterministic**	**Both**
Navy - Numident	44%	18%	38%
Navy - 1940 Census	60%	21%	19%
Numident - 1940 Census	30%	38%	32%
FindAGrave - 1940 Census	24%	34%	42%
Navy - FindAGrave	10%	30%	59%

Finally, I chain links between datasets, using linking pairs with higher information content to infer additional links. This chaining process is particularly useful for inferring links between Navy records and the 1940 Census.

C Additional Results

C.1 Network Formation: Other forces

C.1.1 Empirical Framework

Role of Ship Occupational Similarity:

The second specification focuses on occupational proximity:

$$\begin{aligned}
&\beta_{hd}^{dest} X_{ijkd} = \beta^{dest} Sh_{ijkd} + \beta^{dist} \log \text{dist}_{hd} Sh_{ijkd} \beta^{dest,close} Sh_{ijkd}^{close} + \beta^{dist,close} \log \text{dist}_{hd} Sh_{ijkd}^{close} \\
&\beta_{d}^{home} X_{ijkd} = \beta^{home} Sh_{ijkd} + \beta^{home,close} Sh_{ijkd}^{close}
\end{aligned} \tag{C.1}$$

In this specification, Sh_{ijkd}^{close} represents the share of shipmates from state d who operated in the same rating group. Navy ships were divided into horizontal units (e.g., deck, boiler room), and individuals were more likely to interact frequently with those in their unit. The key parameters $\beta^{dest,close}$, $\beta^{dist,close}$, and $\beta^{home,close}$ capture how these closer interactions influence the strength of network ties in predicting migration.

Role of Socioeconomic Background:

The third specification examines the role of socioeconomic background:

$$\begin{aligned} \beta_{hd}^{dest} X_{ijkd} &= \beta^{dest} Sh_{ijkd} + \beta^{dist} \log \text{dist}_{hd} Sh_{ijkd} \beta^{dest,occ} \text{Occ}_{ijkd} + \beta^{dist,occ} \log \text{dist}_{hd} \text{Occ}_{ijkd} \\ \beta_{d}^{home} X_{ijkd} &= \beta^{home} Sh_{ijkd} + \beta^{home,occ} \text{Occ}_{ijkd} \end{aligned} \tag{C.2}$$

Here, Occ_{ijkd} represents the average imputed occupational score of shipmates from state d. The key parameters $\beta^{dest,occ}$, $\beta^{dist,occ}$, and $\beta^{home,occ}$ capture how interactions with shipmates from higher-income households influence migration decisions. This specification is motivated by the possibility that access to individuals of higher socioeconomic status might provide better information or access to job opportunities in different locations Einiö (2019).

C.1.2 Results

C.1.3 Proximity Effects

Shared shipboard roles, measured by common rating branches, only weakly influence network effects on migration. For most state pairs, the share of shipmates from a given state working in the same rating group does not meaningfully affect migration to that state. However, having more same-rating shipmates from one's home state does increase the likelihood of staying there. This effect persists long-term but is strongest in the short run.

These results suggest that sailors working closely with others from their home state may form tighter social bonds, reinforcing their ties to home. However, measuring these shared experiences presents challenges. Navy ratings change over a sailor's service, and many men spent substantial time in entry-level ratings not clearly tied to specific shipboard divisions. These measurement issues likely contribute to the weak evidence for shared role effects on migration patterns.

C.1.4 Socioeconomic Background Effects

The analysis of socioeconomic background, as measured by average occupational score, reveals no significant impact on migration decisions. This lack of effect is consistent across different specifications and time periods. This finding may reflect the demographic characteristics of the sample: young, non-college-educated men who were not yet established in their careers. In this context, differential exposure to individuals of slightly higher or lower socioeconomic status did not substantially influence migration choices.