NBER WORKING PAPER SERIES

INTERNAL MIGRATION IN THE UNITED STATES: RATES, SELECTION, AND DESTINATION CHOICE, 1850-1940

Ariell Zimran

Working Paper 30384
http://www.nber.org/papers/w30384

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2022

For helpful comments, I thank Jeremy Atack, Brian Beach, William Collins, James Feigenbaum, Sarah Quincy; participants in the Junior Economic History Workshop; and seminar participants at Vanderbilt University. Work on this paper was completed while I was a W. Glenn Campbell and Rita Ricardo-Campbell National Fellow and the William C. Bark National Fellow at the Hoover Institution, Stanford University and while I was a Visiting Scholar at the Opportunity and Inclusive Growth Institute at the Federal Reserve Bank of Minneapolis; funding from both institutions is gratefully acknowledged. The views expressed herein are those of the author and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Internal Migration in the United States: Rates, Selection, and Destination Choice, 1850-1940
Ariell Zimran
NBER Working Paper No. 30384
August 2022
JEL No. J61,N31,N32,O15

ABSTRACT

I study the internal migration of native-born white men in the United States using linked census data covering all possible 10- and 20-year periods 1850--1940. Inter-county migration rates were stable over time. Selection into migration on the basis of occupational status was also largely stable and was neutral or slightly negative. But the orientation of internal migration changed, declining in distance, becoming more directed towards the west, and increasingly driving urbanization. These patterns changed in the 1930s as migration became less common and less urban oriented. These results provide a clearer understanding of historic US internal migration than previously possible.

Ariell Zimran
Department of Economics
Vanderbilt University
2301 Vanderbilt Place
Nashville, TN 37235
and NBER
ariell.zimran@vanderbilt.edu

A data appendix is available at http://www.nber.org/data-appendix/w30384

1 Introduction

Internal migration is fundamental to the American narrative. It has been seen for centuries as a tool for individuals to improve their economic situation (de Tocqueville 1835 [2000]; Turner 1921; Ward 2020), with "great opportunities [lying] just over the horizon" (Brooks 2003). Moreover, by providing an avenue for labor to reallocate to the most productive sectors (Caselli and Coleman 2001; Kuznets 1966), it was a fundamental driver of the United States's transformation into the world's industrial and economic powerhouse.[1] Because of its importance on both the micro and macro scales, internal migration has been the subject of large literatures in economics, economic history, and history, which have provided insights into the nature, causes, and effects of internal migration in specific contexts in US history.[2] But constraints on the data previously available to study internal migration have severely limited scholars' ability to study its long-run patterns, leaving significant and fundamental blind spots in economic historians' understanding of this formative phenomenon in US economic and social history.

This paper documents and describes, for the first time, the rates of, selection into, and destination choice patterns of inter-county migration for US-born white men over the period 1850–1940. My analysis is enabled by recent advances in the availability of complete-count data from the US censuses of this period (Ruggles et al. 2021) and in the technology by which to make links between them (Abramitzky et al. 2021a; Bailey et al. 2020; Zimran 2022a). Building on these advances, I construct 13 datasets linking native-born white men aged 18–40 when first observed over all 10- and 20-year spans in the period 1850–1940. These datasets enable me to overcome the limitations faced in prior studies of the long-run trends in US internal migration: only with linked data is it possible to observe inter-county migration, to separate the flow of migration from its stock, and to measure

[1] See also Lewis (1954). Conversely, Cheremukhin et al. (2017) and Hayashi and Prescott (2008) argue that restrictions on internal migration were responsible for delaying growth in Russia and Japan, respectively.

[2] By this I mean that papers tend to focus on internal migration surrounding a particular event, in a relatively short span of time, or in a specific region within the United States. Studies of specific instances of internal migration in US history include Boustan (2009, 2010), Boustan, Fishback, and Kantor (2010), Boustan, Kahn, and Rhode (2012), Caselli and Coleman (2001), Cheremukhin et al. (2017), Collins and Wanamaker (2014, 2015), Derenoncourt (2022), Eli, Salisbury, and Shertzer (2018), Eriksson and Niemesh (2016), Ferrie (2005), Fishback, Horrace, and Kantor (2006), Fouka, Mazumder, and Tabellini (2022), Hatton and Williamson (1992), Hornbeck (2012, 2020), Hornbeck and Naidu (2014), Long and Siu (2018), Rosenbloom (2002), Salisbury (2014), Sichko (2022), Steckel (1989), Stewart (2006, 2009, 2012), Stuart and Taylor (2021), Vigdor (2002), and Ward (2020).

the selectivity of migration.

I find that the rates of inter-county migration of the native-born adult white male population were remarkably stable between the 1850s and the 1920s at about 33 percent for 10-year spans and about 40 percent for 20-year spans. Selection into migration on the basis of occupational rank was also largely constant over time, with migrants either neutrally or slightly negatively selected.

This constancy contrasts with substantial changes over time in the orientation of internal migration, coming from changes in internal migrants' origins and destination choice patterns. In particular, both the deterrent effect of distance in destination choices and the relative attractiveness of the west increased over the study period. At the same time, the average distance of a move declined and intra-state moves grew to comprise a greater share of inter-county moves, implying that a focus on inter-state moves alone would miss an increasingly large share of migration. Most strikingly, the relationship between internal migration and urbanization changed over my study period. Urbanites were initially more likely than observationally similar ruralists to migrate, but by the twentieth century were either less likely or approximately as likely, depending on the definition of an urban place. The attractiveness of urban areas as destinations for internal migrants also increased over time. The combination of these patterns resulted in a steady increase from the beginning of my study period through the 1920s in the degree to which internal migrants' increase in urbanization over a linkage span exceeded that of stayers. That is, internal migration increasingly became a force driving the urbanization of the economy.

The 1930s marked a change in these patterns in all respects. Selection on the basis of both occupational rank and initial urban status were moderated relative to the earlier twentieth century. More dramatically, a substantial decline in the rates of internal migration occurred for the first time, with 10-year migration rates declining by nearly 8 percentage points, or about 25 percent. This decline was coupled with a substantial retrenchment in the degree to which the urbanization growth or labor demand growth experienced by internal migrants exceeded that of stayers as urban areas became less attractive as destinations. That is, whereas internal migration was a force driving urbanization in the earlier parts of the twentieth century, this was not true in the 1930s.

The contribution of this paper is predicated on the advantages arising from the newfound ability

to make links between all complete-count US censuses 1850–1940. But recent scholarship (e.g., Abramitzky et al. 2021a; Bailey et al. 2020) has brought attention to the danger posed by false links arising from automated linking methods. This is a particularly apposite challenge in studying internal migration because any incorrect match will, in all likelihood, appear as an observation of inter-county migration. This challenge will spuriously increase observed migration rates, conflate selection into migration with selection into false matching, and confound true destination choice patterns with spurious ones generated by false matches. To address this concern, I repeat the main results with alternative matching methods of various strictness and draw only conclusions that are robust to the choice of method. For my estimates of the rates of inter-county migration, which are most sensitive to this danger, I also use information on the ages and birthplaces of children in the household to generate an alternate measure of internal migration (Collins and Zimran 2019; Rosenbloom and Sundstrom 2004). This method does not require linkage but can be applied only to a select sample of individuals and only to study inter-state migration; but it enables me to estimate the rate of false matches for each linkage method and to correct my estimated migration rates. The resulting estimates are largely invariant to the strictness of the linkage method.[3]

The main contribution of this paper is to update, deepen, and expand existing descriptions of US internal migration in the period 1850–1940 (Ferrie 1997a, 2006a,b; Hall and Ruggles 2004; Rosenbloom and Sundstrom 2004).[4] Prior work in this vein has relied on unlinked census data and on information on individuals' state of birth. It has therefore only been able to quantify the stock of inter-state migrants or the rate of inter-state migration of individuals with young children over 10-year spans, and has been extremely limited in its ability to describe the selection and destination choice patterns of migrants. By using linked census data, I am able to provide the first comprehensive description of the rates, selection, and sorting of US inter-county migration. Indeed, given the constraints on prior studies, this is the first description of the rates, selection, and sorting of internal migration flows at any geographic level that is not limited to families with children.[5] On the

[3]Even though the danger of false matches poses a challenge to my results, the advantages of linked data over the previously available data are strong enough that they enable me to answer questions and document facts to an extent not possible with other data sources.

[4]See also Adams and Kasakoff (1985), Oberly (1986), and Villaflor and Sokoloff (1982).

[5]It therefore complements Zimran's (2022c) study of immigrants' secondary migration within the United States.

whole, my findings provide an entirely new view of US internal migration, documenting, for the first time, facts that are interesting and important for their own sake and for the better understanding of US history that they provide. Importantly, the story of internal migration arising from my analysis differs from the one arising from earlier studies of US internal migration over this period—constant rather than declining in frequency before the 1930s and neutrally or slightly negatively rather than positively selected—providing a different interpretation of US internal migration and its interaction with the development of the US economy.[6]

In establishing these facts, this paper adds to several literatures beyond that seeking to describe long-run trends in internal migration. First, it contributes to the large literature studying specific instances of internal migration in the United States, such as frontier migration (Ferrie 1997b; Stewart 2006), the Great Migration of African Americans (Collins and Wanamaker 2014, 2015), and Dust Bowl migration (Hornbeck 2020; Long and Siu 2018; Sichko 2022). Although it has been possible to study these specific instances of migration in detail, the limited understanding of the broad patterns of US internal migration implies that the context into which the findings of this literature fit has not been clear. This paper brings this backdrop into sharper focus. This paper also dovetails with papers describing the rates of and selection into modern internal migration (e.g., Greenwood 1975; Molloy, Smith, and Wozniak 2011). In combination with them, it enables a description of internal migration patterns in the United States over more than 150 years. This paper also relates to recent work taking advantage of the increased ease in making links across complete-count censuses to better describe patterns in mobility in US history, broadly defined to include intergenerational mobility (e.g., Pérez 2019; Ward 2021) and immigrant assimilation (e.g., Abramitzky et al. 2021b; Collins and Zimran 2021; Zimran 2022c) in addition to internal migration. Finally, given the weight assigned to internal migration in the literatures on US economic growth (Kuznets 1966) and on intergenerational mobility (e.g., Ward 2020), the clearer understanding that this paper provides of internal migration adds to these literatures.

Ultimately, this paper addresses basic questions with relatively simple answers. But these questions and the new answers that I provide are fundamental and essential to a complete economic

[6]Moreover, the story that arises from my analysis is deeper than that available in prior research because of the greater ability to study migrant selection and destination choice.

history of internal migration in the United States specifically and to a complete economic history of the United States more generally.

2 Background

The first data collected at a national level with the explicit goal of measuring internal migration were part of the US census of 1940, which inquired as to each individual's place of residence in 1935. According to the Department of Commerce (1939, p. xxiv), the question was added because, "as the rate of population growth ... declined, the factor of internal movements ... [became] increasingly important in determining the future population in the various parts of the country." Questions of this form continued to be included in subsequent censuses, giving scholars the ability to study internal migration from 1935 onwards with relative ease.

Internal migration was, of course, recognized as an important facet of the US economy and culture long before this point (e.g., de Tocqueville 1835 [2000]; Turner 1921). Turner (1921) famously argued that the United States's high internal migration rates in global perspective in the nineteenth century were the product of the availability of land on the frontier, which provided opportunities for surplus labor from eastern cities.[7] Turner (1921) also predicted that these high migration rates would be temporary, declining when the frontier ceased to exist. Others (e.g., Shannon 1945; Weber 1899) challenged claims of uniquely high migration rates in the nineteenth century, arguing that the growth of urban areas would draw migration, counteracting this decline.

Despite this long-running interest in US internal migration, limited data have constrained scholars' ability to actually evaluate these claims. Most of what is known about US internal migration in the pre-war period is based on the census's question on individual's state of birth, which, through the census of 1930, was the only systematically available information on internal migration.[8] The simplest application of these data uses information in census publications to determine the share of the population in each census year that lived outside of the state of birth—that is, the stock of inter-state migrants. Ferrie (1997a, 2006a,b) and Hall and Ruggles (2004) report the results of such

[7] Although these claims were published in book form in 1921, they date to 1893.

[8] This is a surprising paucity of information when compared to the detailed data available on international immigration from as early as the 1820s (Barde, Carter, and Sutch 2006; Ferenczi and Willcox 1929).

an analysis. They find that the stock of white male inter-state migrants was effectively constant (as a share of population) from 1850 to 1940. Some improvement over these simpler tabulations is possible through the use of census microdata. These enable a focus on particular age cohorts, reducing the impact of concerns such as changing age composition.[9] Hall and Ruggles (2004) and Rosenbloom and Sundstrom (2004) do this, finding a strong decline in inter-state migration rates throughout the nineteenth century, with a slight increase from 1900 to 1920 for the young and a continuing decline in this period for the old.[10]

The foremost shortcoming of analyses of this type is that they describe only the stock of inter-state migrants, not the flow, which has long been recognized as problematic: "While the census figures for State of birth form practically the only source of extensive information with regard to interstate movements of population, *extreme care should be exercised in the use of these figures as representing or measuring migration . . . It is therefore with some hesitancy that the term 'migration' is used at all . . .*" (Department of Commerce 1933, p. 135, emphasis added).[11] Relatedly, if an individual moved several times between birth and observation, only one move would be observed. As a result, conclusions drawn from these data regarding migration are largely incomparable to those in, for instance, the literature on the Age of Mass Migration or the literature on modern internal migration (e.g., Molloy, Smith, and Wozniak 2011). An effort to overcome this constraint is made by Rosenbloom and Sundstrom (2004). They use information on the birthplaces and ages of children in order to determine whether their parents had moved over the prior decade, enabling them to observe migration rates rather than stocks. The main limitation of this approach is that it can be applied only to families with young children and can be used only for relatively short spans due to the tendency of children to leave their parents' household around age 18.[12] This analysis yields evidence of a sharp decline in migration rates through the nineteenth century, followed by a

[9] For instance, an increase in life expectancy can cause the stock of internal migrants to increase, since the old would have had a greater time period in which to potentially migrate.

[10] These studies were also limited in that the census microdata available at that time were far less comprehensive than those provided by Ruggles et al. (2021)—they included only samples, not complete-counts, and no data were available for 1930.

[11] A "synthetic cohort" method can observe the net migration flow of a particular group, but not gross migration. Gross migration, however, is the typical interest of the economics of migration.

[12] Collins and Zimran (2019) use a similar approach to divide Irish immigrants into those arriving before or during the Great Irish Famine, but by focusing on the outcomes of children rather than adults do not face the same representativeness issues.

slight increase through the twentieth.

A second limitation arising from these data is that they do not permit the observation of intra-state moves. Such moves are likely to be particularly important in studying rural-to-urban migration (e.g., Department of Commerce 1933, p. 135; Ferrie 2005) and thus to shedding light on the role of such flows in US development.[13]

The available census data also place two main limitations on what can be learned even about the moves that can be observed. Both arise from the fact that individuals are observed only after any migration has taken place. The first consequence of this is that it is not possible to study selection into migration on any but the most basic characteristics, or without implicitly assuming that post-migration education and occupational information reflect an individual's pre-migration characteristics. What information is available has been interpreted by Hall and Ruggles (2004) and Rosenbloom and Sundstrom (2004) to indicate positive selection into inter-state migration throughout US history. The second consequence of these limitations is that the ability to study migrants' destination choices is limited. While individuals' destinations can, of course, be observed, the absence of detailed data on the prior place of residence limits the extent to which the distance of a move can be determined, meaning that it is difficult to determine the drivers of destination choice while accounting for the cost of migration. Moreover, the inability to determine the timing of migration means that conditions in the destination and potential alternative destinations at the time that the move occurred are not known.

All of these limitations can be overcome by using linked census data, which provide an alternative way to measure individuals' internal migration simply by comparing their places of residence in the initial and final census. Such data make it possible to bound the timing of a move, meaning that the flow rather than the stock can be observed, and the finer residence data enable the observation of intra-state inter-county moves.[14] And the pre-migration information enables the direct measurement of migrant selection and the determination of the distance of the move. The main drawback of linked

[13]The fact that birth place is not informative of prior place of residence in the United States for the foreign born also implies that it is not possible to observe the internal migration of immigrants. For this paper, the omission of the foreign born is not consequential because of the focus on the native born. But Zimran (2022c) takes advantage of the fact that linked data enable the measurement of immigrants' internal migration to study immigrant distribution during the Age of Mass Migration.

[14]In principle, it is possible to observe migration at an even finer geographic level.

data is that there are certain segments of the population that it is systematically impossible to link, most notably women. But for white men, it is, in principle, possible to link everyone, meaning that migration can be observed without limiting the sample to men with children.

Several studies have exploited such data to study internal migration.[15] Steckel (1988, 1989), for instance, used data on census records linked between the census of 1850 and 1860 to analyze patterns of westward migration.[16] A number of other studies (e.g., Collins and Wanamaker 2014, 2015; Ferrie 1999; Long and Siu 2018; Stewart 2006, 2009) have also studied the rates, selection, and sorting of specific instances of internal migration with linked data. But until recently, it was not possible to construct linked datasets with sufficient coverage to reveal the broad long-term patterns of US internal migration. As a result, there is no systematic study of inter-county migration in the United States for the period before 1935, and even what is known about inter-state migration is extremely limited as a result of the constraints described above.

Recent advances in data availability and record linkage methods have unlocked the potential of these data, enabling me to improve on the limitations of existing research. In particular, fully digitized data on the characteristics of every individual in every census have only recently become available, enabling for the first time systematic linkage of the white male population as a whole from 1850–1940.

2.1 Modern Internal Migration

In contrast to the very limited picture that exists of internal migration before 1935, internal migration since then is extremely well documented due both to the prior-place-of-residence question in the census and to the availability of more detailed data, such as the CPS, the ACS, and tax records, some of which directly link individuals' residences over time. As in the historical literature, there is a substantial body of work studying specific instances of internal migration in the United States

[15]The earliest iteration of this approach are the so-called "community studies," which infer migration through the failure to locate individuals observed in a particular location in one census in that same location in subsequent censuses (e.g., Coleman 1962; Curti 1959; Galenson and Pope 1989; Malin 1935; Throne 1959).

[16]Steckel (1989) did rely on the birthplaces of children in order to assist in manual linkage of individuals to their records in the earlier census.

or using internal migration as a setting in which to study other questions.[17]

As in the historical case, there is also a literature describing the basic characteristics of US internal migration, though the more comprehensive data in this regard imply that measurement is far more straightforward. The broad findings in this regard in terms of rates, selection, and sorting are summarized by Molloy, Smith, and Wozniak (2011). The modern literature largely begins around 1980 and shows evidence of a decline in internal migration rates since then. Combined with results from Ferrie (2006a,b), Hall and Ruggles (2004), and Rosenbloom and Sundstrom (2004), which describe the intervening decades, the postwar picture of internal migration is of an increase until about 1980 followed by a decrease since then. There is also evidence of positive selection into internal migration (Wozniak 2010). By creating the first comprehensive series of migration rates (rather than stocks) with known timing for the period before 1940, this paper enables, for the first time, the dovetailing of modern and historical internal migration rate series, and therefore the construction, for the first time, of a series of internal migration rates spanning over 150 years of US history.

3 Data

My analysis is based on 13 datasets making all possible 10- and 20-year links between the US censuses of 1850–1940. I begin the analysis in 1850 because that year's census was the first to enumerate the entire free population. The analysis ends in 1940 because, at the time of writing, this is the most recent complete-count census that has been fully digitized. I created these datasets by merging complete-count census records provided by Ruggles et al. (2021) with the linkage crosswalks provided by Zimran (2022a).[18] These datasets provide information on an individual's county of residence in each of the two censuses, which enables me to determine whether an individual made an inter-county move between them.[19]

[17] See, for instance, Amior (2021), Conway and Rork (2016), Huang, South, and Spring (2017), Huttunen, Møen, and Salvanes (2018), Jauer et al. (2019), Kennan and Walker (2011), Molloy, Smith, and Wozniak (2017), Oswald (2019), and Treyz et al. (1993).

[18] I use the "basic" links from this set of crosswalks.

[19] To address changing county boundaries, I use county boundary shapefiles from Manson et al. (2019) to define a move to have occurred if an individual is observed living in a final-year county that does not overlap geographically with his initial-year county. An example of how this definition is applied is given in Online Appendix Figure A.1.

Throughout my analysis, I restrict attention to native-born white men aged 18–40 in the initial census. The restriction to men is motivated by an inability to consistently link women between censuses due to name changes at marriage. This limitation is the main way in which linked census data are fundamentally constrained in their ability to reveal internal migration. The restriction to those aged 18–40 in the initial census is intended to ensure that men are observed in the labor force in the initial census of the linkage span while ensuring that they are also not so old that mortality is an important concern. The limitation to whites is made because this is the only group whose internal migration can be tracked beginning in 1850 (because of the omission of most of the black population from the census in 1850 and 1860) and because Collins and Wanamaker (2014, 2015) have already described the internal migration of blacks in detail. Finally, the limitation to the native born is made because Zimran (2022c) has already studied the internal migration of immigrants.

There are two main concerns that arise in the use of linked census data to study internal migration. The first is the danger of false matches—that is, the concern that the linked datasets may not actually describe the same person in the two census years. This concern has been highlighted recently by Abramitzky et al. (2021a) and Bailey et al. (2020), and touches all aspects of the analysis of internal migration.[20] Since nearly all false matches are to individuals living in a different county,[21] the observed rate of migration conflates true migration with false matches. Selection into migration is also, therefore, conflated with selection into false matching. Finally, under the assumption that a false match links an individual in one census to a random individual in a subsequent one, true destination choice patterns are conflated with a tendency for false matches to show spurious migration towards more populous areas.

I address the danger posed by false matches in two ways. First, I repeat all of my analysis using data sets constructed by four alternate linkage methods. Two of these methods are simply different in their linkage parameters—the ABE-NYSIIS and ABE-Exact conservative methods using linkage crosswalks provided by Abramitzky et al. (2020). The other two methods are stricter, reducing the

[20]The concern is also different than in studies of specific cases of internal migration. For instance, in these studies, the goal is often not measurement but the establishment of causal relationships. In these cases, the presence of linkage errors attenuates estimates, strengthening conclusions that certain relationships existed. In this paper, the goal is one of measurement, meaning that the challenge coming from false matches is more fundamental.

[21]This figure can be directly computed, as explained below.

danger of false matches at the potential risk of a less representative sample. The first such method makes a match only when the Zimran (2022a) method and the two ABE methods agree;[22] I refer to this as the *intersection-of-matches method*. The second is even stricter, using only the subset of the intersection of matches in which the match is corroborated by all ostensibly time-invariant information not used in the linkage.[23] I call this the *intersection-of-matches plus corroboration* method. To ensure that my results are not the product of false matches, I draw only conclusions that are robust to the choice of linkage method. In studying the rates of internal migration, I also use an additional method, described in detail below, to estimate the rate of false matches for each linkage method and to correct my estimated internal migration rates for them.

The second main concern in the use of linked data is that they may not be representative of all individuals at risk for linkage.[24] Indeed, this is one reason why, even though it is tempting to simply use the strictest linkage method to minimize the danger of false matches, I do not do so—this would increase the danger of constructing an unrepresentative sample. To address this concern, I reweight each linked dataset so that its observables match (as closely as possible) the distribution of observables of those at risk for linkage in the initial census.[25]

Besides information on an individual's initial place of residence, the initial census of each span provides data on a variety of individuals' pre-migration characteristics, including occupation, literacy, and initial urban status, which I use to study migrant selection. Literacy, though flawed and potentially changing in definition over time, is the only consistently available measure of human capital. Occupation is the only measure of economic status that is available in a consistent way over the complete 1850–1940 period. I use Ruggles et al.'s (2021) occupational codes to construct a measure of occupational rank as follows. First, following Collins and Zimran (2021), I assign to each

[22] Abramitzky et al. (2021a) show that the intersections of several sets of matches have a lower false positive rate than individual methods.

[23] For example, I require, where parents' birthplaces are provided in both censuses, that they agree across the sources.

[24] Linkage rates are presented in Online Appendix Table A.1. Online Appendix Figure A.2 compares the observable characteristics of the linked sample and the sample at risk for linkage for each linkage span.

[25] To reweight the data, I estimate, for each census linkage span, probit regressions for the probability of successful linkage as a function of all observables in the initial census. I then reweight the linked data by the inverse of the estimated conditional linkage probability. In addition to the data in the census, I also use information on name length and commonality from Zimran (2022b). As is standard in such settings, this approach cannot address selection into linkage on the basis of unobservable characteristics.

individual an occupational score based either on the average wealth holdings of each occupation in 1870 or based on the average income of the occupation in 1900 (Preston and Haines 1991). Then, based on these two scores and the occscore variable provided by Ruggles et al. (2021), which is in turn based on data from the 1950 census, I determine the rank of each occupation relative to the white male population aged 18–74 in each census.[26] My occupational status measure is the simple average of these three ranks. To determine whether an individual resided in an urban area in the initial and final census of each linkage span, I use the official census bureau definition of an urban place as one with at least 2,500 inhabitants, as well as an alternative definition using a population cutoff of 25,000.[27]

Finally, I construct a number of measures of the characteristics of an individual's move and of his counties of initial and final residence. These include the distance of the move and whether it crossed state or regional (i.e., census divisions) boundaries; the share of each county's population residing in urban areas under various definitions;[28] and the Bartik (1991)-type labor demand growth experienced by a county. The latter is computed using national-level employment growth rates by industry and a potential destination county's initial-year employment-by-industry composition. I address changing boundaries using Hornbeck's (2010) method.

4 Inter-county migration rates, 1850–1930

I begin by answering the most basic, but perhaps the most fundamental question about US internal migration—what was the rate of inter-county migration? That is, how likely was it that an individual living in a county in a given census year would move to a different county over the next 10 or 20 years? In this section, I present results using all five of the linkage methods introduced above in order to illustrate the potential error arising from errors in linkage and how these errors can be

[26] Following Collins and Zimran (2021), I probabilistically reclassify the occupations of men in agriculture in 1850 who lived with a head of household who was a farmer and to whom they were related. This addresses the changing definition of a farmer from that census to future ones. I also follow Collins and Zimran (2021) in assigning family members of farmers the average of the occupational status of farmers and farm laborers. I also use crosswalks from Zimran (2022b) to check the robustness of my results to assigning imputed occupations to individuals whose occupation code in the Ruggles et al. (2021) data is listed as "Not Yet Classified." These results are presented in Online Appendix B.

[27] I also show results defining an urban area as one of the top 10 cities in the United States.

[28] I also compute population density.

addressed.

Figure 1 presents the estimated uncorrected rates of inter-county migration over 10- and 20-year spans. The estimated 10-year inter-county migration rates by my main linkage method are 43.6 percent in the 1850s, rise slightly to 45 percent in the 1860s, and then decline steadily to 37.8 percent by the 1920s. For 20-year spans, this linkage method indicates a decline in migration rates from 52.4 percent in the 1850–1870 span to 46.7 percent in the 1910–1930 period before declining more sharply to 42.9 percent for the 1920–1940 period. Besides these results, four patterns are immediately clear. First, the estimated 20-year inter-county migration rates are in all cases greater than the corresponding (i.e., same starting year and linkage method) 10-year estimates. Second, for every 10-year link and for every 20-year link other than the 1850–1870 link by the more permissive methods, more than half of individuals are found living in the same county in both censuses. These patterns point to the general validity of the linkage methods—higher 20-year than 10-year migration rates are to be expected, and finding a majority of individuals in the same county after 10 years could not occur without some validity in linkage.[29]

The third pattern evident in these results is that, for each method and span length, declines in inter-county migration rates are evident over time, with somewhat larger declines among the less restrictive methods than the more restrictive ones. Finally, the three more permissive linkage methods point to higher migration rates than do the two more restrictive linkage methods. These two patterns are strongly indicative of the estimates being affected by errors in linking—both the decreasing migration rates over time and the increasing migration rate with the declining strictness of matches are to be expected if false matches drive false observations of migration and if the probability of a false match declines over time.

Fortunately, as discussed above, it is also possible to measure internal migration without relying on linkage by using the birthplaces and ages of children in the household—though this can be done only for families with children and only for inter-state migration—to compare the rates of inter-state migration implied by household structure and by linkage in order to estimate the rate of false linkage between censuses, and subsequently to use these estimated false match rates to correct the

[29]If all matches were simply random, the share of stayers would be below one percent.

estimated inter-county migration rates for the whole sample.

I apply this intuition as follows.[30] First, for individuals in the linked sample whose household structure enables it, I use the birthplace and age composition of children in their households to determine who did and did not migrate over the previous 10 years. Movers are defined as those with a child aged less than 10 years old born in a different state than the current state of residence, and no older child born in the state of residence. Stayers are defined as those with a child born in the state of residence at least 10 years old and no children younger than 10 born in a different state. This categorization is performed for the latter census of each 10-year span (i.e., in 1860 for the 1850–1860 span) for the linked sample; it is not applied to 20-year spans because children are unlikely to remain in their parents' household for over 20 years. This procedure results in a subsample of the linked dataset composed of individuals whose inter-state migration status over a 10-year span can be measured in two ways—by comparing their state of residence in the initial and final censuses and according to their household composition.[31]

According to the law of total probability, the probability that an individual in this linked and categorized sample is observed to have made an inter-state move (whether he truly moved or not) according to his residence state in the initial and final census, which I denote as $P(\text{moved state})$, can be written as

$$P(\text{moved state}) = P(\text{moved state}|\text{true match})[1 - P(\text{false match})] + P(\text{moved state}|\text{false match})P(\text{false match}). \quad (1)$$

Rearranging equation (1), I can express the probability of a false match as

$$P(\text{false match}) = \frac{P(\text{moved state}) - P(\text{moved state}|\text{true match})}{P(\text{moved state}|\text{false match}) - P(\text{moved state}|\text{true match})}. \quad (2)$$

To calculate equation (2), I use the following quantities from the sample of individuals who were linked and whose migration status could be determined by the household-composition method. As

[30] Zimran (2022c) also applies this method.

[31] As with the definition of inter-county migration, I define an individual as having made an inter-state move only if his final-year and initial-year states of residence share no geographic overlap.

P(moved state), I use the estimated rate of inter-state migration according to the comparison of the initial- and final-year residence state. As P(moved state|true match)—the true probability of an inter-state move—I use the estimated rate of inter-state migration according to the household-composition method. Finally, as P(moved state|false match)—the probability of observing inter-state migration in the case of a false match—I use an individual's birthplace and age to determine the average probability that a person to whom he could be linked would live in a different state. For instance, a 20-year old Alabama-born man in the 1850 census can be linked, according to my benchmark linkage method, to any Alabama-born man aged 26–34 in the 1860 census.[32] To determine the probability that a false match would create an observation of inter-state migration for this man, I determine the fraction of Alabama-born 26–34-year old men in 1860 who lived in a different state than the man originally lived in in 1850. Performing analogous calculations for all men in the sample and averaging the results yields my estimate of P(moved state|false match).

Figure 2 presents the estimated rates of false linkage for each method and initial census year. The estimates presented in Figure 2 fit with expectations. The estimated false match rates are higher for the more permissive linkage methods and decline over time from about 15 percent to under 10 percent by the end of the study period. For the more restrictive methods, the estimated false match rates are initially about 5 percent and fall to approximately zero in the 20th century.[33]

Having computed an estimate of P(false match) based on inter-state migration in the subset of the linked sample for which the household-composition method can also be applied, I can now use this estimate to correct my estimates of inter-county migration for the full linked sample. To do this, I rearrange equation (1) and replace state moves with county moves to yield

$$P(\text{moved county}|\text{true match}) = \frac{P(\text{moved county}) - P(\text{moved county}|\text{false match})P(\text{false match})}{1 - P(\text{false match})} \tag{3}$$

This is an expression for an estimate of inter-county migration rates that is both uninfluenced by false matches and can be computed using the available data and estimates. Specifically, to compute this

[32] All of the other methods that I use permit only a gap of two years or less in the age-implied birthyear, and so the approach is adjusted in those cases to accommodate this alternate requirement.

[33] Imprecision in some cases results in estimates that are below zero by a very small margin. The largest in magnitude is of -0.4 percent for spans beginning in 1920 by the intersection-of-matches plus corroboration method.

value I use the following quantities. As P(false match), I use the estimate computed from equation (2) above; for 10-year spans, I use the estimate from the analogous span; for 20-year spans, I use the estimate for the 10-year span beginning in the same year; there is no 1880–1890 span, and so I must omit the 1880–1900 span. As P(moved county), I use the probability of observing an inter-county move in the full linked dataset. For inter-county migration, P(moved county|false match) is sufficiently close to one that,[34] with minimal loss, I can write equation (3) as

$$P(\text{moved county}|\text{true match}) = \frac{P(\text{moved county}) - P(\text{false match})}{1 - P(\text{false match})}. \tag{4}$$

The estimates coming from computing equation (4) are my benchmark estimates of inter-county migration rates that are corrected for false matches. Despite my efforts to correct these estimated migration rates, it is inevitable that there will remain some error. There are many possible causes. One in particular is that, as discussed above, the subsample for which inter-state migration can be observed by both linkage and household composition is not representative of the broader linked sample. This is a somewhat less severe issue than when using the household composition method to directly measure migration, since the internal migration of men with children is likely very different from that of all men; in this case an issue arises only insofar as the probability of a false match differs between these groups. This method also relies on the assumption that the true migration rates of the correctly linked and incorrectly linked are the same.[35] Since it is well established that selection into linkage is non-random, it is not likely that this assumption holds in reality—characteristics determining success in linkage likely also determine the propensity to move. The method also assumes that there is no error (on average) in determining individuals' migration status using the household-composition method. This method is also not informative as to *which* observations are incorrectly linked, since it is used simply to deflate an aggregate quantity.[36] This implies that it

[34]It is, in principle, possible to compute this figure as above for inter-state migration, but in practice, the estimates are very close to one. In Online Appendix C, however, where I focus on moves of at least 150 miles, P(moved county|false match) is no longer approximately 1, and I therefore estimate it following the method described there.

[35]This is because the correction operates, in essence, by applying the migration rate of the correctly linked portion of the population to match the full population.

[36]For this reason, I do not study conditional changes in internal migration rates over time, as, for instance, Rosenbloom and Sundstrom (2004) do.

cannot be used in the more detailed analyses below of selection and destination choice. Despite these limitations, however, and as discussed above, the use of linked data to study US internal migration provides important advances over existing methods. The strategy that I have used to correct for linkage error, while necessarily imperfect, provides the best possible glimpse into the history of US internal migration.

Figure 3 presents the estimated corrected migration rates for each linkage method and span. For both 10- and 20-year spans, the resulting predictions of inter-county migration rates and their changes over time are similar across the linkage methods.[37] For 10-year spans, the estimated migration rates by my preferred linkage method are consistently between 31.0 and 35.3 percent from the 1850–1860 span to the 1920–1930 span with no clear trend. I discuss the 1930s in more detail below. The 20-year migration rates are, as expected, higher, at 39.6 to 43.6 percent and evolve largely without trend save for a slight decline between the two spans of the nineteenth century. The main takeaway is that the frequency of inter-county migration over the period 1850–1930 was largely constant.

5 Migrant selection, 1850–1930

Characterizing migrant selection is crucial to a basic description of any flow of migration. Doing so also provides insight into the potential effects of internal migration on individuals, on the economies of sending and receiving areas, and on the broader economy. By providing information on the pre-migration characteristics of prospective migrants, linked data enable me to delve further into this question than has previously been possible. However, it remains possible that errors in linkage may influence conclusions if the probability of false matches varies along a dimension on which I study selection. I present here results using only my preferred linkage method, and limit the conclusions that I draw to those that are robust to the use of alternate and stricter linkage methods (results in Online Appendix D).

[37] It is notable that the estimates, even for the less restrictive methods, are similar to those arrived at with the stricter matching methods without correction, suggesting that the danger of unrepresentativeness due to strict matching is minimal. It is, of course, still non-zero. The remaining differences between methods after correction could be due to selection of less mobile types into the stricter linked sample.

I focus first on two measures of migrant selection—occupational rank and literacy in the initial year of the span—that speak to whether internal migrants were positively or negatively selected. These correspond roughly to the typical focus in the economics of migration on earnings (e.g., Borjas 1987; Chiquiar and Hanson 2005) and education (e.g., Card 2005). To measure selection, I estimate, separately for each linked dataset, the equation

$$y_{it} = \alpha_t + \beta_t r_{it} + \gamma_t \ell_{it} + \delta_t u_{it} + \mathbf{x}'_{it}\varphi + \varepsilon_{it}, \tag{5}$$

where y_{it} is an indicator equal to one if individual i migrated across county lines in span t, r_{it} is individual i's initial-year occupational rank, ℓ_{it} is individual i's initial-year literacy, u_{it} is individual i's initial-year urban residence, and $\mathbf{x}_{it}$ is a vector of initial-year controls.[38] The β_t, γ_t, and δ_t coefficients provide measures of migrant selection in span t, indicating whether greater occupational rank, literacy, or urban status were associated with a greater probability of migration. To ensure that changes in the β_t, γ_t, and δ_t coefficients across censuses are the product of actual changes in selection patterns rather than of changing availability of controls in the census, and to avoid confounding the interpretation of the coefficients,[39] I limit the vector $\mathbf{x}_{it}$ to the variables listed in Online Appendix Table A.2 that are available in all initial censuses.[40]

Results on migrant selection on the basis of occupational rank and literacy are presented in Figure 4. Each panel contains two sets of estimates—one with no controls and one with controls, including initial-county fixed effects and the other dimensions of selection.[41] Panels (a) and (b) focus on selection on the basis of occupational rank. Unconditional selection into inter-county migration was consistently negative, moving from very close to zero in the beginning of the study period to more strongly negative in the later nineteenth century and in the twentieth.[42] Such unconditional

[38]Note that, despite the time subscripts in equation (5), this is not a panel regression. Instead, the subscripts indicate that the equation is estimated separately for each linkage span, yielding separate coefficient estimates for each.

[39]For instance, I do not wish to include both an indicator for farm residence and for urban residence, as these contain largely the same information. Similarly, I do not include multiple measures of occupational status.

[40]I also do not include every observable in Online Appendix Table A.2, as some of these might confound the interpretation of others (e.g., I do not include more than one measure of occupational status). Ultimately, the controls are age, urban residence, literacy, occupational rank, household size, and whether the individual was a head of household.

[41]These results are based on estimating equation (5) with u_{it} defined based on the 2,500-person cutoff.

[42]These results are largely insensitive, even in magnitude, to the strictness of the linkage method, meaning that

comparisons, however, ignore differences in age, sector, distance to potential destinations, and other factors likely to influence migration. When conditioning on all observables, including initial county of residence, selection patterns are much more stable over time. For spans beginning in 1850, conditional selection on the basis of occupational rank was zero or positive. A notable decline into spans beginning in 1860 then occurred, indicating more negative selection into migration in this span than earlier. From this point through the 1920s, the coefficients hover around -0.05. With the occupational rank measure ranging from zero to one and having a standard deviation of roughly 0.25 throughout the study period, these coefficients are small but not negligible: a one-standard deviation increase in occupational rank was associated with about a 1.25-percentage point decline in migration probability on a base ranging from 38 to 52 percent.[43] The variation in these coefficients over time is even smaller. For 10-year spans, the variation is of no more than 0.043 (which corresponds to about a one-percentage point change in migration probability for a standard-deviation change in rank), whereas for 20-year spans the variation is of no more than 0.011 (corresponding to 0.03-percentage point change in migration probability for a standard-deviation change in rank). In sum, beginning in the 1860s, individuals of greater occupational rank were somewhat less likely or as likely to make an inter-county move than otherwise similar individuals from the same initial county but of lower occupational rank,[44] and this difference remained largely constant over time.

Panels (c) and (d) of Figure 4 focus on selection into internal migration on the basis of literacy. All estimates indicate that literate individuals were less or as likely as otherwise similar individuals to make an inter-county move (including those for alternate linkage methods). The precise patterns diverge between 10- and 20-year spans. For 10-year spans, a decline in the magnitude of negative selection from 1850–1860 to 1860–1870 is followed by near total constancy in selection. For 20-year spans, the decline in the magnitude of selection persisted throughout the nineteenth century before stabilizing in the twentieth. The coefficients are again small but non-negligible, with literate individuals about 2 to 4 percentage points less likely to migrate than illiterates.

linkage concerns are not a major issue.

[43] The appropriate reference point is the uncorrected migration rate by Zimran's (2022a) linkage method since the data used for the analysis are not corrected for false matches. The more negative coefficient over time is inconsistent with the results being driven by false matches, since these became less common over time, as shown above.

[44] That is, individuals of greater occupational rank were less likely to move according to the main estimates, but stricter linkage methods show a coefficient closer to zero.

On the whole, these results paint a picture of migration that was neutrally or slightly negatively selected, with this selection, by most indications—including on the basis of the best available measure of socioeconomic status—largely constant from the 1860s through the 1920s.

6 Changes in the origin and orientation of internal migration, 1850–1930

In this section, I study selection into internal migration on the basis of initial urban residence as well as the destination choices of internal migrants. I find that the results of general constancy in the rates and selection of internal migration conceal important changes in the nature of US internal migration. Again, I present results here for my main linkage method with results by alternate methods presented in Online Appendix D.

Figure 5 focuses on selection into inter-county migration on the basis of urban residence in the initial year of each span using two different measures of urbanization—the 2,500-inhabitant cutoff (panels a and b) and an indicator for being in a city of at least 25,000 (panels c and d).[45] As above, for each measure and span, I include two sets of estimates—an unconditional estimate and one that conditions on all controls, but I use state fixed effects instead of county fixed effects.[46] In spans beginning in 1850, urbanites were 5 to 10 percentage points more likely to migrate than otherwise similar ruralists, depending on the linkage method. This conditional urban migration premium declined over the nineteenth century. By spans beginning in 1880 for 20-year spans or in 1900 for 10-year spans, this pattern reversed, with urbanites approximately as likely (for the 2,500-person definition) or up to 5 percentage points less likely (for the 25,000-person definition) than otherwise similar ruralists from the same state to migrate. Ruralists' conditional migration premium then declined over the twentieth century. Thus, in contrast to selection on the basis of occupational rank and literacy, selection on the basis of urban residence was relatively larger in magnitude, changed throughout the 1850–1930 period, and for larger cities changed in sign such that urbanites were

[45] Results for population density and urban residence defined as the 20 largest cities in the country are presented in Online Appendix Figure A.3.

[46] I do not include county fixed effects because many of these measures vary only across counties. Including county fixed effects would lead identification to come from only a small number of observations.

initially more and then less likely than ruralists to move.[47]

Changes over time are also evident in internal migrants' destination choice patterns. Figure 6 focuses on the distance distribution of moves. Panels (a) and (b) divide moves into inter-county but intra-state, inter-state but intra-region, and inter-region. This method of division is particularly important as the fraction of moves that are intra-state is a measure of the extent to which internal mobility is not observed when using state of birth to determine migration. Inter-county but intra-state moves were an important component of US internal migration, accounting for about 40 percent of moves at the beginning of the study period and rising over time to about 60 percent by the end of the study period. This increase came at the expense of inter-region moves, with the share of inter-state but intra-region moves remaining largely constant in frequency over time. Panels (c) and (d) present violin plots for distance of moves (i.e., box plots with the complete distribution overlaid). They show that the distribution of move distance had a consistent peak below 50 miles. In the earlier periods, a second peak around 500 miles is also evident and fades over time, as the upper tail thins.

To what extent were these unconditional changes the product of changing individual characteristics over time? Using a pooled dataset of all migrants from all linkage spans of a particular length, I estimate a regression of the form

$$y_{it} = \beta_t + \mathbf{x}'_{it}\delta + \varepsilon_{it}, \tag{6}$$

where y_{it} is some measure of individual i's move distance in span t, β_t is a series of span fixed effects, and $\mathbf{x}_{it}$ is a set of initial-year controls, including all observables available in all census years, and either initial-year state or initial-year county fixed effects. Figure 7 plots the β_t resulting from estimating equation (6) with no controls, with controls and initial-state fixed effects, and with controls and initial-county fixed effects. Omitting the indicator for spans beginning 1850, these estimates show how migration distance changed over time. Paralleling Figure 6, all measures show

[47]There are also changes over time in migrants' origins. In Online Appendix Figures A.4 and A.5, I compute migration rates for individuals living in each county at the start of each 10- or 20-year span. These maps omit any county for which the sample contains fewer than 30 observations. The high-emigration area moves westward over time. The Northeast consistently shows low rates of emigration, whereas migration rates from the South and the West were consistently relatively high. The Midwest shows a change over time, however, relatively high in the nineteenth century but relatively low by the end of the study period.

a large unconditional decline in the distance of the move or in the probability of moving across states or regions over time. The decline, however, was moderated by changes in the demographics of migrants, meaning, for instance, that the move made by an individual in 1850–1860 was about 100 miles further than that made by an observationally similar person in 1920–1930.

Determining which characteristics of destinations attracted migrants is less straightforward, as simply comparing destination characteristics over time is confounded by changes in the economy.[48] Instead, I restrict attention to movers and estimate a conditional logit model of the form

$$P_{iotj} = \frac{\exp(\beta_t d_{oj} + \mathbf{z}'_{jt}\delta_t)}{\sum_k \exp(\beta_t d_{ok} + \mathbf{z}'_{kt}\delta_t)},$$

where P_{iotj} is the probability that migrant i initially from county o in linkage span t chose destination j, $\mathbf{z}_{jt}$ is a vector of characteristics of county j in the initial year of span t, d_{oj} is the distance between counties o and j, and k indexes all potential destinations. The vector $\mathbf{z}$ includes a county's initial-year urbanization (i.e., the fraction of the population living in urban areas) or Bartik (1991)-type labor demand growth, and census-division fixed effects (with New England as the excluded category).[49] The coefficients β_t quantify the deterrent effect of distance in span t, and the coefficients δ_t quantify the attractiveness of the various county characteristics in that period. The coefficients have the usual interpretation of coefficients in a logit model—the marginal effect of the variable in question on the log odds of selecting a particular destination, conditional on its distance.[50] In principle, estimating this model is computationally difficult because of the large number of options for individuals to choose from (all counties in the United States other than county o).[51] Fortunately, Guimarães, Figueirdo, and Woodward (2003) provide a method to simplify this estimation, making it tractable, though still computationally costly.[52] The cost of this simplification is that I am

[48] For instance, the average urbanization of destinations might increase over time simply because of the urbanization of the economy.

[49] As with my study of selection, I limit the contents of the vector $\mathbf{z}$ to avoid confounding the interpretation of the variables that I do include.

[50] Since this method only makes within-span comparisons, it is not confounded by changes in the characteristics of destinations over time.

[51] Any final-year county with borders overlapping those of county o must also be excluded because of how inter-county migration is defined.

[52] Guimarães, Figueirdo, and Woodward (2003) do not use weights. I adapt their method to the use of weights by replacing the number of individuals choosing a given option with the sum of normalized weights of individuals choosing a particular destination. The method is computationally costly because it requires the inclusion of origin-county fixed

not able to control for different characteristics of internal migrants. Instead, the estimates must be thought of as providing a measure of the attractiveness of various location characteristics to internal migrants as a whole over time.[53]

The results of this estimation are presented in Figure 8. Panels (a)–(d) focus on the deterrent effect of distance and on the attractiveness of urban areas.[54] An increase over time in the deterrent effect of distance is evident from the increasing magnitude of its negative coefficient over time. This result is consistent with the declining distance of moves described above. Conversely, an increase over time in the attractiveness of urban destinations is clear through the 1920s. The change in the magnitude of this coefficient is large. In the nineteenth century, a completely urban county had a 1.5 greater log odds of being selected than an entirely rural destination.[55] By the 1920s, this figure had risen to about 3 or 3.5, depending on the definition of an urban place.[56]

Panels (e) and (f) of Figure 8 compare the attractiveness of each census division as a destination relative to New England. For the most part, there is little change over time. The main exceptions are the Mountain and Pacific divisions, however, which increased substantially in their relative attractiveness, indicating that there was some characteristic of these areas other than their urbanization that increasingly attracted internal migrants.[57]

Finally, I compare the growth over the linkage span of the urbanization or labor demand growth of movers' and stayers' residence counties by estimating an equation of the form

$$y_{it_2} - y_{it_1} = \alpha_t + \beta_t m_{it} + \mathbf{x}'_{it}\gamma_t + \varepsilon_{it},$$

where y_{it_2} is the value for individual i's residence county in the final year of span t, y_{it_1} is the residence county in the initial year, m_{it} is an indicator equal to one if individual i moved during

effects. I cluster standard errors by destination county for conservative inference.

[53] In principle, it is possible to control for these characteristics by dividing the sample along any relevant characteristic and estimating the model separately, but the estimation burden is then effectively doubled for each binary variable added.

[54] I focus attention here on urban areas defined as places with a population of at least 2,500 or 25,000. Results for other measures of urban status, as well as population density and labor demand growth are presented in Online Appendix Figure A.6.

[55] This corresponds to an increase in the odds by a factor of about 4.5.

[56] This corresponds to a 20- to 30-fold increase in the odds.

[57] Online Appendix Figure A.7 presents the same results controlling for labor demand growth instead of urbanization.

span t, and $\mathbf{x}_{it}$ is a vector of individual i's observables in the initial year of span t, including indicators for initial county of residence.[58] For stayers, the y_{it} values describe the same county in both periods, meaning that stayers capture the change in the urbanization of their home county. A positive β_t implies that movers experienced a greater gain in urbanization than stayers from the same county—that is, that they moved to areas growing more quickly.

Results are presented in Figure 9.[59] In essence, these results combine the selection and sorting patterns described above, in which urbanites became relatively less likely to migrate as compared to ruralists and urban areas increased in attractiveness over time. In the nineteenth century, I find that movers and stayers experienced approximately the same increase in urban status and labor demand growth. By the twentieth century, these coefficients had shifted to become strongly positive. Thus, whereas nineteenth-century migration was relatively neutral with respect to urbanization, twentieth-century migration was a force driving the contemporaneous urbanization of the economy by shifting population into more urban (or at least more rapidly urbanizing) areas.

7 The 1930s

All of these patterns changed in the 1930s. For the first time, this decade was marked by a substantial decline in the rate of internal migration (Figure 3 panel a), which fell from 34.2 percent for 1920–1930 to 27.5 percent for 1930–1940 according to my preferred linkage method. The 1930s also marked a moderation in selection into migration on the basis of occupational rank (Figure 4 panel a). Relative to the 1920s, there was a considerable increase in the coefficient that was larger than its change in any other period in the twentieth century. This change marked either a moderation of negative selection or a transition from somewhat negative to somewhat positive selection depending on the linkage method; that is, internal migrants were less negatively selected in the 1930s than in the 1920s. A similar, though somewhat weaker, moderation is present in terms of selection into internal migration on the basis of urban residence in that urbanites and ruralists were most similar in terms of migration propensity than ever before in the 20th century, though this was the product of

[58] This is computed following Bartik (1991) as the product of initial-year employment shares and national growth between the two census years.

[59] Results for population density and the country's 20 largest cities are presented in Online Appendix Figure A.8.

a trend that evolved throughout the 20th century (Figure 5 panels a and c). Similarly, the distance of moves was lowest in the 1930s, again continuing a declining trend (Figure 7 panels a, c, and e).

A particularly striking change concerns the attractiveness of urban areas as destinations of internal migrants. From the 1850s to the 1920s, the attractiveness of urban areas as destinations for internal migrants increased. But the 1930s marked a reversal of this trend, with a decline in this measure to levels approximately equal to those of the 1910s (Figure 8 panels a and c). Even more dramatically, the 1930s marked a sharp reversal in the degree to which the urbanization growth or labor demand growth of movers exceeded that of stayers (Figure 9 panels a, c, and e). This figure had climbed from the 1850s to the 1920s, but reversed sharply in the 1930s. For instance, in the 1920s, movers' growth in the likelihood of living in a city of 25,000 or more inhabitants was about 8 percentage points greater than that of stayers. By the 1930s this figure had fallen to zero—a level not seen since the late nineteenth century—implying that the urbanization growth of movers and stayers was nearly identical and that internal migration was no longer a force driving urbanization.

8 Summary of Robustness Checks

In addition to verifying, in Online Appendix D, that the conclusions that I draw are robust to the choice of linkage method, I also verify the robustness of my results to a number of other permutations of the sample or definitions of variables. In Online Appendix C, I redefine migration such that an individual must both cross county lines and move at least 150 miles to be considered an internal migrant. Naturally, the estimated migration rates are lower, in the vicinity of 15 percent over 10-year spans and 20 percent over 20-year spans. There is also a slight decline (about 2 to 3 percentage points) in these rates in the nineteenth century, which on the whole results in a slight downward trend, even when omitting the 1930s. Selection into migration was also different, with urbanites more likely than ruralists to make longer-distance moves. Other results are largely unaffected.

In Online Appendix B, I repeat the main results using imputed occupational codes from Zimran (2022b) in cases where the occupations given by Ruggles et al. (2021) are listed as "Not Yet Classified." In Online Appendix E, I omit from the sample individuals with foreign-born fathers. The intention of this exercise is to ensure that the continued assimilation of second-generation im-

migrants does not affect the results. Zimran (2022c) shows that immigrants' internal migration patterns differed from those of natives. This exercise reduces the influence of these patterns.

9 Discussion

My results bring internal migration in the United States over the period 1850–1940 into clearer focus than has previously been possible. In summary, I find that inter-county migration was largely constant in its frequency and selection, but declined in distance over time, became increasingly attracted to the west, and changed from a force that was relatively neutral in its effects on urbanization and drew more from urban than from rural areas into one that was a force towards the urbanization of the US economy, and which drew more (or at least not less) from rural than from urban areas. The 1930s then marked a change in all regards, with migration declining in frequency, becoming somewhat more neutrally selected, and sharply declining in the degree to which it drove increases in urbanization.

These findings deepen our understanding of US internal migration. Regarding the rates of internal migration, existing research (Ferrie 1997a; Hall and Ruggles 2004; Rosenbloom and Sundstrom 2004) has identified a decline in inter-state migration over the period 1850–1940, though the studies differ on the precise timing of this decline. My findings for inter-county migration paint a different picture—one of stable migration rates until the 1930s. In part, this difference in results can be attributed to the greater ability of linked data to bound the timing of migration. But it can also be attributed to my finding that inter-county but intra-state migration became increasingly important as average migration distances declined, meaning that a focus solely on inter-state migration would overlook an increasingly large share of internal population movements over time.[60]

My results also give an understanding of the selection and sorting of internal migration that goes into far greater depth than has previously been possible. There are few existing estimates of migrant selection over a broad span to which my results can be compared. But the estimates that do exist (Hall and Ruggles 2004; Rosenbloom and Sundstrom 2004) point to positive selection—something

[60]Consistent with the decline being driven by not observing shorter-distance moves, Online Appendix C shows that there *is* a decline in migration rates when limiting migration to moves of at least 150 miles.

for which I find no evidence in my analysis. The existing descriptions of migrants' destination choices in the long run (Hall and Ruggles 2004; Rosenbloom and Sundstrom 2004) are even more limited, generally focusing on the region of destination or the urban or rural status of the destination. My analysis provides a richer description than has previously been possible.

The results of this paper also help to better understand the development of the US economy over the period that I study, and in particular shed new light on Turner's (1921) interpretation of internal migration in US history and later critiques of this interpretation. Consistent with Turner (1921), I find evidence that urbanites were more likely to internally migrate in the nineteenth century. But I find no evidence of strongly negative selection into internal migration,[61] nor do my conditional logit results indicate that rural areas were particularly attractive in the nineteenth century (i.e., the coefficient on urban is always positive).

I also find no evidence of a decline in migration rates after the closing of the frontier, contrary to findings for inter-state migration. Instead, I find that the closing of the frontier was followed by a twentieth century marked by a shift of internal migration into a force driving the urbanization of the economy from the perspective of both selection and destination choice. Beyond validating critiques of Turner's (1921) claims that the nineteenth century was exceptional in its high rates of internal migration, tThis finding speaks to the structural transformation of the economy as it shifted from agricultural to industrial (Caselli and Coleman 2001; Kuznets 1966), with internal migration helping to allocate labor to the nation's growing industrial sector in urban areas, as evidenced by my results regarding the labor demand growth experienced by movers. My results also shed light on the unique nature of internal migration in the 1930s. Whereas this decade is popularly viewed as one of higher mobility spurred by the Depression and Dust Bowl, my results show that migration rates were, in fact, lower during this period (see also Long and Siu 2018).

Finally, the results of this paper also introduce or deepen a number of puzzles. The first concerns the relatively constant frequency of internal migration over the first 80 of the 90 years that I study. The structural transformation of the US economy, combined with dramatic changes in transportation technology, land availability, international immigration, and labor market integration from 1850 to

[61] Of course, this does not consider selection on unobservables (e.g., Ferrie 1997a; Stewart 2006).

1930, would be expected to lead to some change the frequency or selectivity in internal migration. Indeed, there is evidence that these forces were associated with substantial changes in immigrant assimilation (Collins and Zimran 2021) and intergenerational mobility (e.g., Long and Ferrie 2013; c.f., Ward 2021). Yet evidently such a change in the frequency of internal migration did not occur until the basket of shocks of the 1930s. A similar pattern is evident in the trend over census years in the total stock of lifetime interstate migrants. This paper cements this puzzle as applying to the flow of migration rather than only to its stock. Another puzzle concerns the increase over time in the deterrent effect of distance in the destination choices of internal migrants. This increase came despite substantial improvements over the study period in transportation technology. This specific result can be rationalized by observing that the share of migrants who moved shorter distances increased,[62] but a larger question concerns why the increase in short-distance migration occurred. Explaining either of these puzzles is beyond the scope of this paper, but my documentation of them lays them out as the targets of future research.

10 Conclusion

Internal migration is one of the fundamental forces that contributed to the development of the American economy and identity. But a lack of suitable data has made it difficult to establish even the most basic facts of internal migration in US history. In this paper, I exploit recent advances in the availability of complete-count census data and in the technology to make links between censuses in order, for the first time, to describe the trends in the rates, selection, and sorting of the inter-county migration of native-born white men over the period 1850–1940. I find that the rates of and selection into migration were generally constant over time, with largely neutral or slightly negative selection into migration. But the distance of moves declined over time, migrants became increasingly attracted to the west, and the origins and orientation of migration shifted considerably over this period to become increasingly oriented to driving a flow of population toward urban areas. These findings are simple, but they are novel and foundational to a complete economic history of the United States.

[62] In this sense, it is not strictly correct to call this a deterrent *effect* of distance.

References

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

Abramitzky, Ran, Leah Platt Boustan, Katherine Eriksson, Santiago Pérez, and Myera Rashid (2020). *Census Linking Project: Version 2.0 [dataset]*. `https://censuslinkingproject.org`.

Abramitzky, Ran, Leah Platt Boustan, Elisa Jácome, and Santiago Pérez (2021b). "Intergenerational Mobility of Immigrants in the US Over Two Centuries." *American Economic Review* 111:2, pp. 580–608.

Adams, John W. and Alice Bee Kasakoff (1985). "Wealth and Migration in Massachusetts and Maine: 1771–1798." *Journal of Economic History* 45:2, pp. 363–368.

Amior, Michael (2021). "The Contribution of Immigration to Local Labor Market Adjustment." CEP Discussion Paper 1678.

Bailey, Martha, Connor Cole, Morgan Henderson, and Catherine Massey (2020). "How Well Do Automated Methods Perform in Historical Samples? Evidence from New Ground Truth." *Journal of Economic Literature* 58:4, pp. 997–1044.

Barde, Robert, Susan B. Carter, and Richard Sutch (2006). "Table Ad106–120: Immigrants, by country of last residence—Europe, 1820–1997." In *Historical Statistics of the United States*. Susan B. Carter, Scott Sigmund Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright (ed.). Cambridge: Cambridge University Press, pp. 1.560–1.563.

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

Borjas, George J. (1987). "Self-Selection and the Earnings of Immigrants." *American Economic Review* 77:4, pp. 531–553.

Boustan, Leah Platt (2009). "Competition in the Promised Land: Black Migration and Racial Wage Convergence in the North, 1940–1970." *Journal of Economic History* 69:3, pp. 756–783.

——— (2010). "Was Postwar Suburbanization 'White Flight'? Evidence from the Black Migration." *Quarterly Journal of Economics* 125:1, pp. 417–443.

Boustan, Leah Platt, Price V. Fishback, and Shawn Kantor (2010). "The Effect of Internal Migration on Local Labor Markets: American Cities during the Great Depression." *Journal of Labor Economics* 28:4, pp. 719–746.

Boustan, Leah Platt, Matthew E. Kahn, and Paul W. Rhode (2012). "Moving to Higher Ground: Migration Response to Natural Disasters in the Early Twentieth Century." *American Economic Review: Papers and Proceedings* 102:3, pp. 238–244.

Brooks, David (2003). "The Triumph of Hope Over Self-Interest." *New York Times* CLII, January 12, 2003, WK15.

Card, David (2005). "Is the New Immigration Really So Bad?" *Economic Journal* 115, F300–F323.

Caselli, Francesco and Wilbur John Coleman (2001). "The US Structural Transformation and Regional Convergence: A Reinterpretation." *Journal of Political Economy* 109:3, pp. 584–616.

Cheremukhin, Anton, Mikhail Golosov, Sergei Guriev, and Aleh Tsyvinsky (2017). "The Industrialization and Economic Development of Russia through the Lens of a Neoclassical Growth Model." *Review of Economic Studies* 84, pp. 613–649.

Chiquiar, Daniel and Gordon H. Hanson (2005). "International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States." *Journal of Political Economy* 113:2, pp. 239–281.

Coleman, Peter J. (1962). "Restless Grant County: Americans on the Move." *Wisconsin Magazine of History* 46, pp. 16–20.

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, pp. 220–252.

——— (2015). "The Great Migration in Black and White: New Evidence on the Selection and Sorting of Southern Migrants." *Journal of Economic History* 75:4, pp. 947–992.

Collins, William J. and Ariell Zimran (2019). "The Economic Assimilation of Irish Famine Migrants to the United States." *Explorations in Economic History* 74, 101302.

——— (2021). "Working Their Way Up? US Immigrants' Changing Labor Market Assimilation in the Age of Mass Migration." NBER Working Paper 26414.

Conway, Karen Smith and Jonathan C. Rork (2016). "How Has Elderly Migration Changed in the Twenty-First Century? What the Data Can–and Cannot–Tell Us." *Demography* 53:4, pp. 1011–1025.

Curti, Merle E. (1959). *The Making of an American Community: A Case Study of Democracy in a Frontier Community*. Stanford: Stanford University Press.

de Tocqueville, Alexis (1835 [2000]). *Democracy in America*. Harvey C. Mansfield and Delba Winthrop (ed.). Chicago: University of Chicago Press.

Department of Commerce (1933). *Fifteenth Census of the United States: 1930, Population*. Vol. II. Washington: Government Printing Office.

——— (1939). *Twenty-Seventh Annual Report of the Secretary of Commerce*. Washington: Government Printing Office.

Derenoncourt, Ellora (2022). "Can You Move to Opportunity? Evidence from the Great Migration." *American Economic Review* 112:2, pp. 369–408.

Eli, Shari, Laura Salisbury, and Allison Shertzer (2018). "Ideology and Migration after the Civil War." *Journal of Economic History* 78:3, pp. 822–861.

Eriksson, Katherine and Gregory Niemesh (2016). "Death in the Promised Land: The Great Migration and Black Infant Mortality." Mimeo., Miami University.

Ferenczi, Imre and Walter F. Willcox (1929). *International Migrations*. New York: National Bureau of Economic Research.

Ferrie, Joseph P. (1997a). "Long-Term Perspectives on Geographic and Economic Mobility, 1800–1900." Mimeo., Northwestern University.

——— (1997b). "Migration to the Frontier in Mid-Nineteenth Century America: A Re-Examination of Turner's 'Safety Valve'." Mimeo., Northwestern University.

——— (1999). *Yankeys Now: Immigrants in the Antebellum United States, 1840–1860*. New York: Oxford University Press.

——— (2005). "How Ya Gonna Keep 'Em Down on the Farm [When They've Seen Schenectady]? Rural-to-Urban Migration in 19th Century America, 1850–70." Mimeo., Northwestern University.

——— (2006a). "Internal Migration." In *Historical Statistics of the United States*. Susan B. Carter, Scott Sigmund Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright (ed.). Cambridge: Cambridge University Press, pp. 1.489–1.494.

——— (2006b). "Table Ac1–42: Interstate migration—native-born population, by race and residence within or outside the state of birth: 1850–1990." In *Historical Statistics of the United States*. Susan B. Carter, Scott Sigmund Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright (ed.). Cambridge: Cambridge University Press, pp. 1.495–1.497.

Fishback, Price V., William C. Horrace, and Shawn Kantor (2006). "The Impact of New Deal Expenditures on Mobility During the Great Depression." *Explorations in Economic History* 43, pp. 179–222.

Fouka, Vasiliki, Soumyajit Mazumder, and Marco Tabellini (2022). "From Immigrants to Americans: Race and Assimilation During the Great Migration." *Review of Economic Studies* 89:2, pp. 811–842.

Galenson, David W. and Clayne L. Pope (1989). "Economic and Geographic Mobility on the Farming Frontier: Evidence from Appanoose County, Iowa, 1850–1870." *Journal of Economic History* 49:3, pp. 635–655.

Greenwood, Michael J. (1975). "Research on Internal Migration in the United States: A Survey." *Journal of Economic Literature* 13:2, pp. 397–433.

Guimarães, Paulo, Octávio Figueirdo, and Douglas Woodward (2003). "A Tractable Approach to the Firm Location Decision Problem." *Review of Economics and Statistics* 85:1, pp. 201–204.

Hall, Patricia Kelly and Steven Ruggles (2004). "'Restless in the Midst of Their Prosperity': New Evidence on the Internal Migration of Americans, 1850–2000." *Journal of American History* 91:3, pp. 829–846.

Hatton, Timothy J. and Jeffrey G. Williamson (1992). "What Explains Wage Gaps Between Farm and City? Exploring the Todaro Model with American Evidence, 1890–1941." *Economic Development and Cultural Change* 40:2, pp. 267–294.

Hayashi, Fumio and Edward C. Prescott (2008). "The Depressing Effect of Agricultural Institutions of the Prewar Japanese Economy." *Journal of Political Economy* 116:4, pp. 573–632.

Hornbeck, Richard (2010). "Barbed Wire: Property Rights and Agricultural Development." *Quarterly Journal of Economics* 125, pp. 767–810.

——— (2012). "The Enduring Impact of the American Dust Bowl: Short- and Long-run Adjustments to Environmental Catastrophe." *American Economic Review* 102:4, pp. 1477–1507.

——— (2020). "Dust Bowl Migrants: Identifying an Archetype." NBER Working Paper 27656.

Hornbeck, Richard and Suresh Naidu (2014). "When the Levee Breaks: Black Migration and Economic Development in the American South." *American Economic Review* 104:3, pp. 963–990.

Huang, Ying, Scott J. South, and Amy Spring (2017). "Racial Differences in Neighborhood Attainment: The Contributions of Interneigborhood MIgration and In Situ Change." *Demography* 54:5, pp. 1819–1843.

Huttunen, Kristiina, Jarle Møen, and Kjell G. Salvanes (2018). "Job Loss and Regional Mobility." *Journal of Labor Economics* 36:2, pp. 479–509.

Jauer, Julia, Thomas Liebig, John P. Martin, and Patrick A. Puhani (2019). "Migration as an Adjustment Mechanism in the Crisis? A Comparison of Europe and the United States 2006–2016." *Journal of Population Economics* 32:1, pp. 1–22.

Kennan, John and James R. Walker (2011). "The Effect of Expected Income on Individual Migration Decisions." *Econometrica* 79:1, pp. 211–251.

Kuznets, Simon (1966). *Modern Economic Growth: Rate, Structure, and Spread*. New Haven: Yale University Press.

Lewis, Arthur W. (1954). "Economic Development with Unlimited Supplies of Labor." *The Manchester School* 22:2, pp. 139–191.

Long, Jason and Joseph P. Ferrie (2013). "Intergenerational Occupational Mobility in Great Britain and the United States since 1850." *American Economic Review* 103:4, pp. 1109–1137.

Long, Jason and Henry Siu (2018). "Refugees from Dust and Shrinking Land: Tracking the Dust Bowl Migrants." *Journal of Economic History* 78:4, pp. 1001–1033.

Malin, James C. (1935). "The Turnover of Farm Population in Kansas." *Kansas Historical Quarterly* 4, pp. 339–372.

Manson, Steven, Jonathan Schroeder, David Van Riper, and Steven Ruggles (2019). *IPUMS National Historical Geographic Information System: Version 14.0* [database]. Minneapolis: IPUMS.

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

——— (2017). "Job Changing and the Decline in Long-Distance Migration in the United States." *Demography* 54:2, pp. 631–653.

Oberly, James W. (1986). "Westward Who? Estiamtes of Native White Interstate Migration After the War of 1812." *Journal of Economic History* 46:2, pp. 431–440.

Oswald, Florian (2019). "The Effect of Homeownership on the Option Value of Regional Migration." *Quantitative Economics* 10:4, pp. 1453–1493.

Pérez, Santiago (2019). "Intergenerational Occupational Mobility Across Three Continents." *Journal of Economic History* 79:2, pp. 383–416.

Preston, Samuel H. and Michael R. Haines (1991). *Fatal Years: Child Mortality in Late Nineteenth Century America*. Princeton: Princeton University Press.

Rosenbloom, Joshua L. (2002). *Looking for Work, Searching for Workers: American Labor Markets during Industrialization*. New York: Cambridge University Press.

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." *Research in Economic History* 22, pp. 289–325.

Ruggles, Steven, Sarah Flood, Sophia Foster, Ronald Goeken, Jose Pacas, Megan Schouweiler, and Matthew Sobek (2021). *Integrated Public Use Microdata Series: Version 11.0* [dataset]. Minneapolis: University of Minnesota.

Salisbury, Laura (2014). "Selective Migration, Wages, and Occupational Mobility in Nineteenth Century America." *Explorations in Economic History* 53, pp. 40–63.

Shannon, Fred A. (1945). "A Post-Mortem on the Labor-Safety-Valve Theory." *Agricultural History* 19, pp. 31–37.

Sichko, Christopher (2022). "Migrant Selection and Sorting during the Great American Drought." Mimeo., USDA ERS.

Steckel, Richard H. (1988). "Census Matching and Migration: A Research Strategy." *Historical Methods* 21:2, pp. 52–60.

——— (1989). "Household Migration and Rural Settlement in the United States, 1850–1860." *Explorations in Economic History* 26, pp. 190–218.

Stewart, James I. (2006). "Migration to the Agricultural Frontier and Wealth Accumulation, 1860–1870." *Explorations in Economic History* 43, pp. 547–577.

——— (2009). "Economic Opportunity or Hardship? The Causes of Geographic Mobility on the Agricultural Frontier, 1860–1880." *Journal of Economic History* 69:1, pp. 238–268.

——— (2012). "Migration to US Frontier Cities and Job Opportunity, 1860–1880." *Explorations in Economic History* 49, pp. 528–542.

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, pp. 134–175.

Throne, Mildred (1959). "A Population Study of an Iowa County in 1950." *Iowa Journal of History* 57, pp. 305–330.

Treyz, George I., Dan S. Rickman, Gary L. Hunt, and Michael T. Greenwood (1993). "The Dynamics of US Internal Migration." *Review of Economics and Statistics* 75:2, pp. 209–214.

Turner, Frederick Jackson (1921). *The Frontier in American History*. New York: Henry Holt and Company.

Vigdor, Jacob L. (2002). "The Pursuit of Opportunity: Explaining Selective Black Migration." *Journal of Urban Economics* 51:3, pp. 391–417.

Villaflor, Georgia C. and Kenneth L. Sokoloff (1982). "Migration in Colonial America: Evidence from Militia Muster Rolls." *Social Science History* 6:4, pp. 539–570.

Ward, Zachary (2020). "Internal Migration, Education and Upward Rank Mobility: Evidence from American History." *Journal of Human Resources* Forthcoming.

——— (2021). "Intergenerational Mobility in American History: Accounting for Race and Measurement Error." NBER Working Paper 29256.

Weber, Adna Ferrin (1899). *The Growth of Cities in the Nineteenth Century: A Study in Statistics*. New York: MacMillan.

Wozniak, Abigail (2010). "Are College Graduates More Responsive to Distant Labor Market Opportunities?" *Journal of Human Resources* 45:4, pp. 944–970.

Zimran, Ariell (2022a). *Census Linkage Crosswalks for US Censuses 1850–1940*. Ann Arbor: Inter-university Consortium for Political and Social Research [distributor]. doi:10.3886/E171701V3.

——— (2022b). *Miscellaneous US Census Crosswalks, 1850–1930*. Ann Arbor: Inter-university Consortium for Political and Social Research [distributor]. doi:10.3886/E171861V2.

——— (2022c). "US Immigrants' Secondary Migration and Geographic Assimilation during the Age of Mass Migration." *Explorations in Economic History* 85, 101457.

Figures

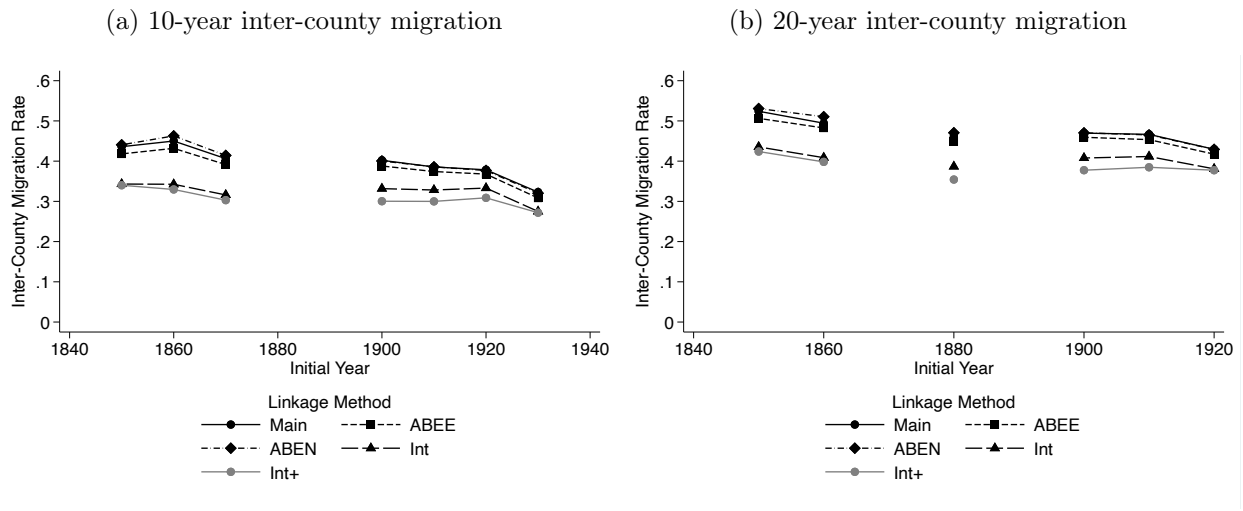


Figure 1: Inter-county migration rates by linkage method and span

Note: Each figure shows the probability that an individual in the linked sample beginning in the year on the x-axis was observed living in a different county 10 or 20 years later according to each linkage method. All observations are weighted by inverse linkage probability.

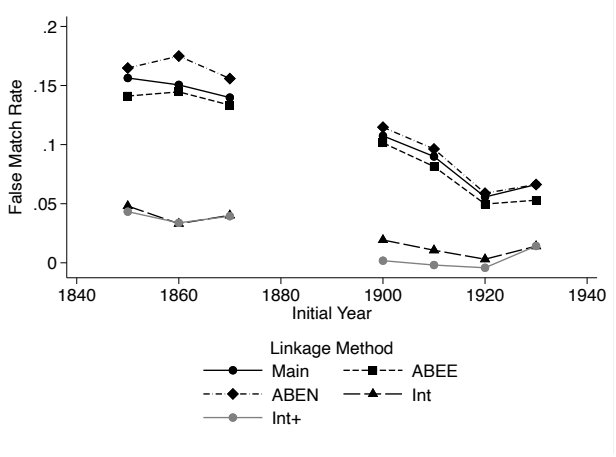
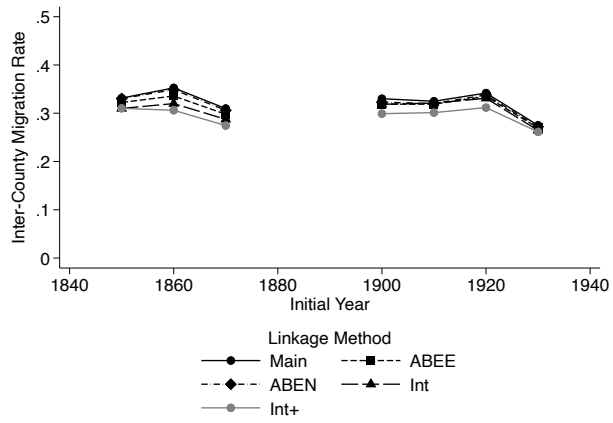


Figure 2: Estimated rates of false linkage

Note: This figure shows the estimated rate of false matching for each linkage method and 10-year span according to the comparison of inter-state migration estimates by the linkage method and by the method based on the birth places and ages of children. The rate is computed according to Bayes's Theorem, as explained in text.

(a) 10-year inter-county migration



(b) 20-year inter-county migration

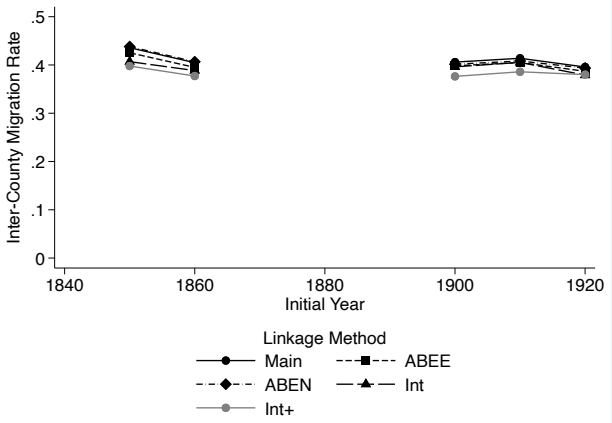
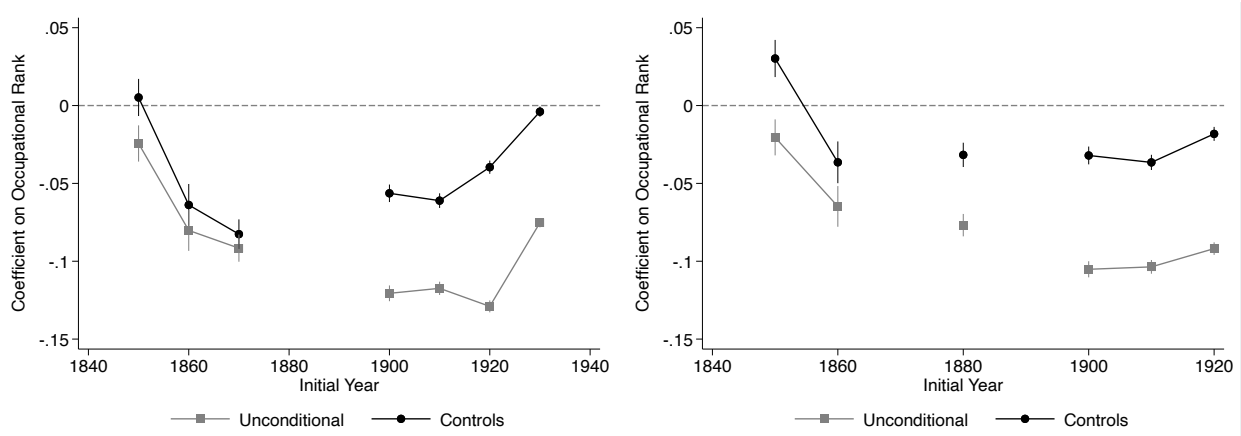


Figure 3: Inter-county migration rates by linkage method and span, corrected for false matches

Note: Each figure shows the probability that an individual in the linked sample beginning in the year on the x-axis was observed living in a different non-overlapping county 10 or 20 years later, according to each linkage method. All observations are weighted by inverse linkage probability. The estimates are then adjusted for false matches according to the method presented in text.

(a) Occupational rank, 10-year spans

(b) Occupational rank, 20-year spans



(c) Literacy, 10-year spans

(d) Literacy, 20-year spans

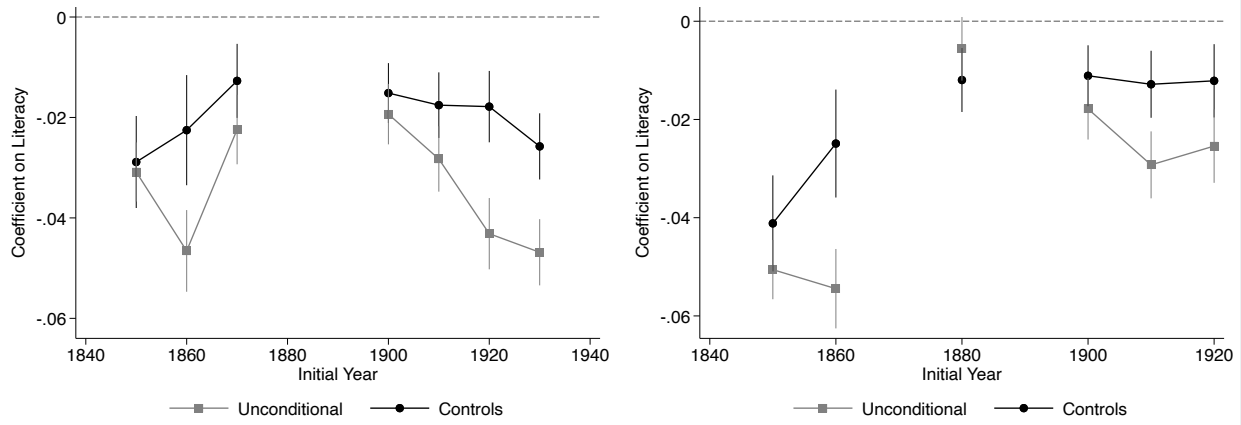


Figure 4: Migrant selection by measure and span, individual characteristics

Note: Each figure shows the coefficient on the variable in question in a selection regression for migration over the stated span. All observations are weighted by inverse linkage probability. Vertical lines indicate 95-percent confidence intervals. Controls are all controls available in all census years, as well as county fixed effects.

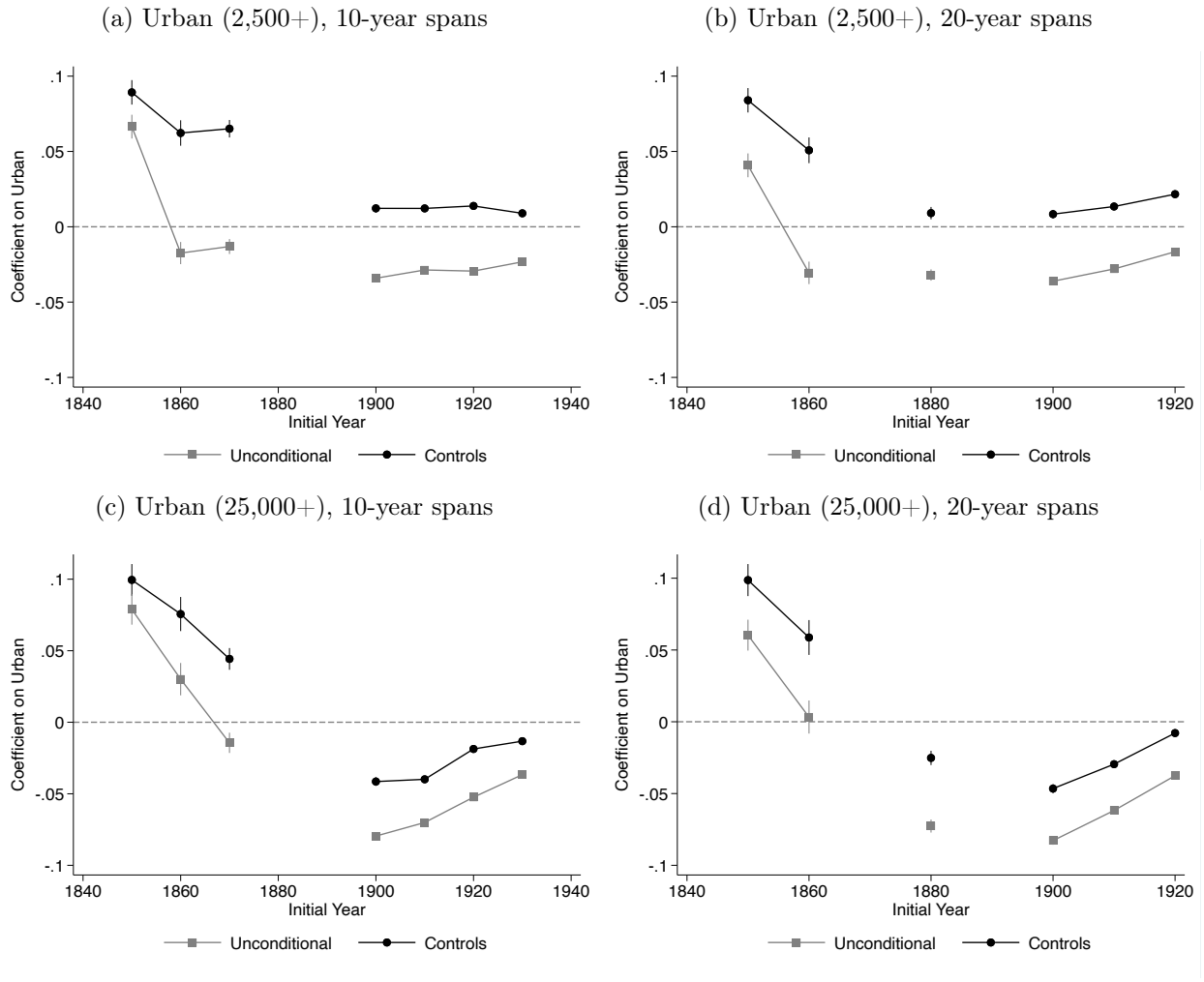
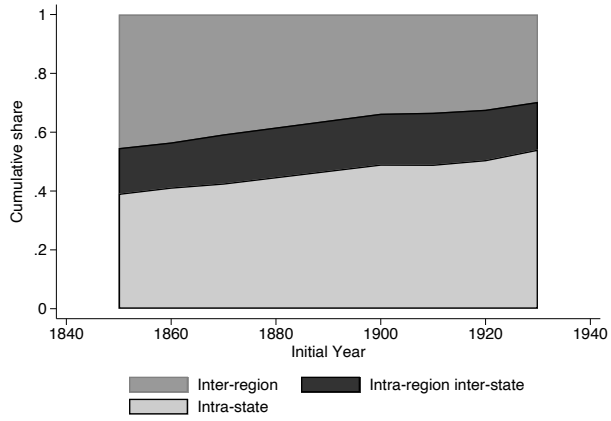


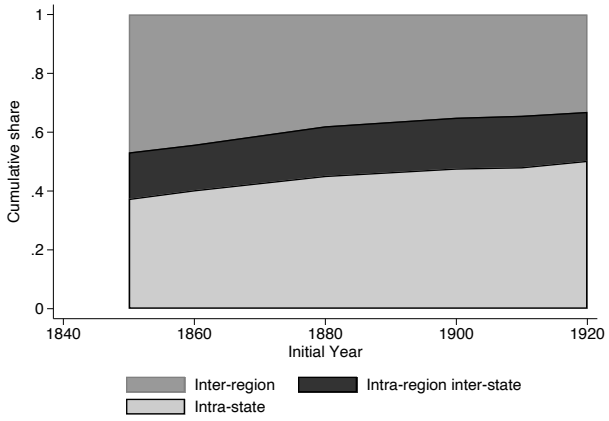
Figure 5: Migrant selection by measure and span, urban residence

Note: Each figure shows the coefficient on the variable in question in a selection regression for migration over the stated span. All observations are weighted by inverse linkage probability. Vertical lines indicate 95-percent confidence intervals. Controls are all controls available in all census years, as well as state fixed effects.

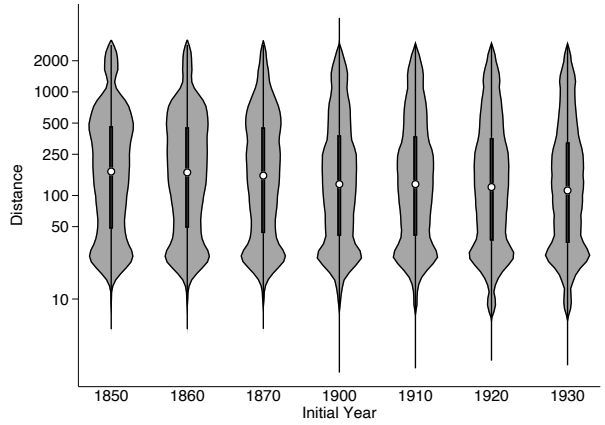
(a) Interstate move, 10-year spans



(b) Interstate move, 20-year spans



(c) Distance of move (miles), 10-year spans



(d) Distance of move (miles), 20-year spans

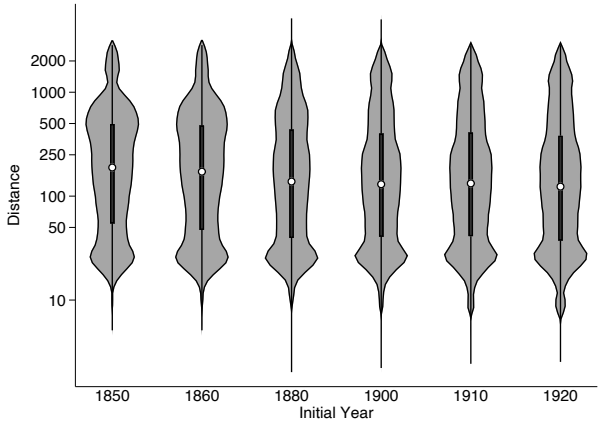
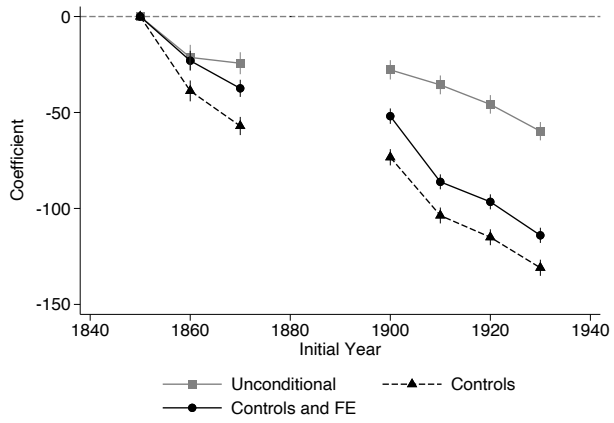


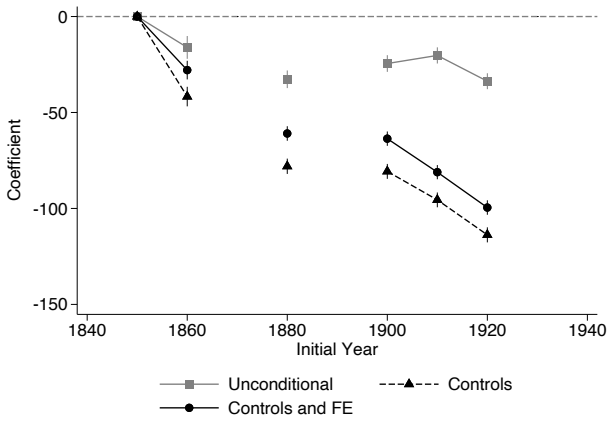
Figure 6: Distance of moves

Note: Panels (a) and (b) show the cumulative distribution of intra-state, inter-state but intra-region, and inter-region moves. Panels (c) and (d) present the distribution of move distance for each linkage span with box plots presenting the mean and quartiles.

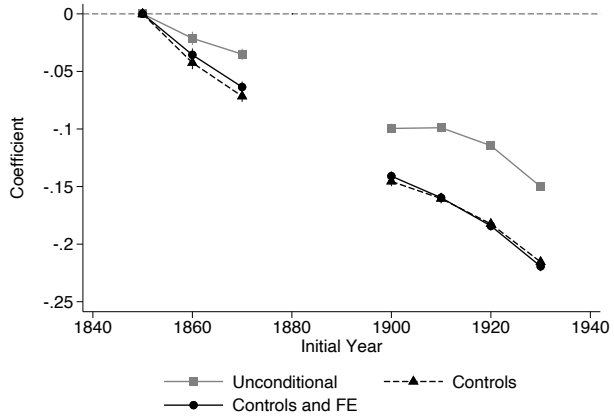
(a) Distance of move (miles), 10-year spans



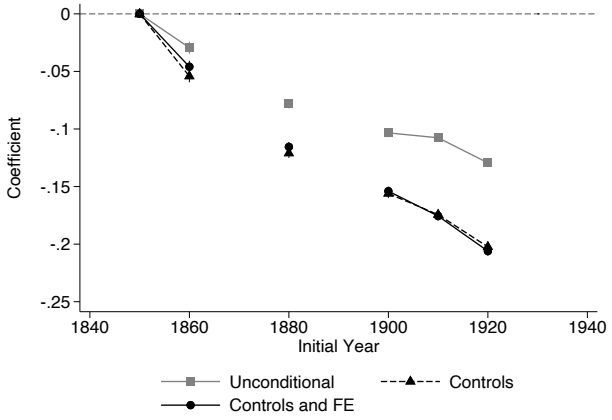
(b) Distance of move (miles), 20-year spans



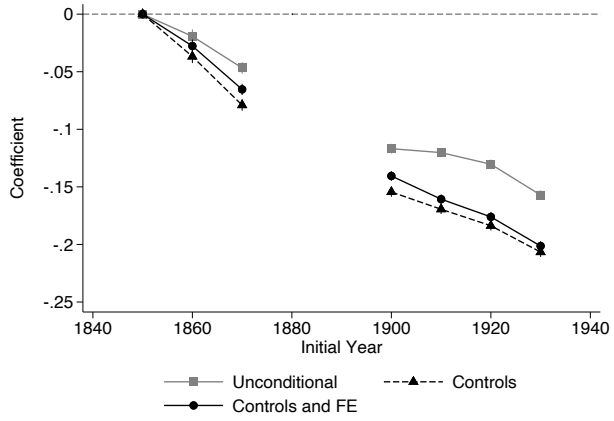
(c) Interstate move, 10-year spans



(d) Interstate move, 20-year spans



(e) Interregion move, 10-year spans



(f) Interregion move, 20-year spans

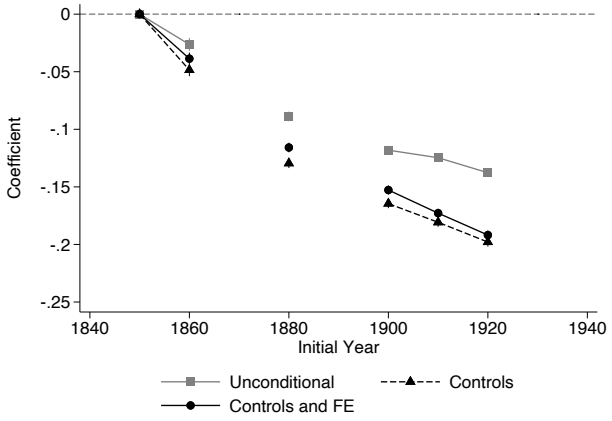
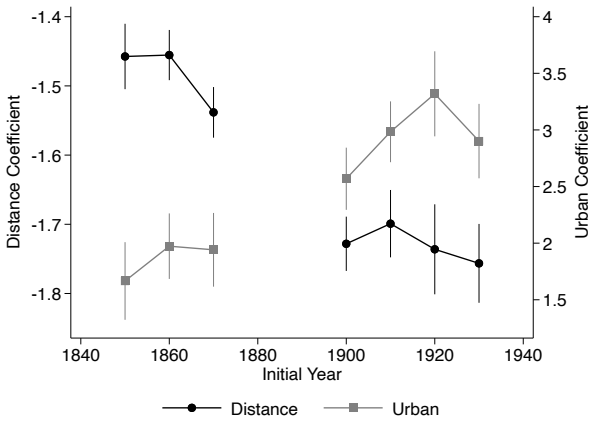


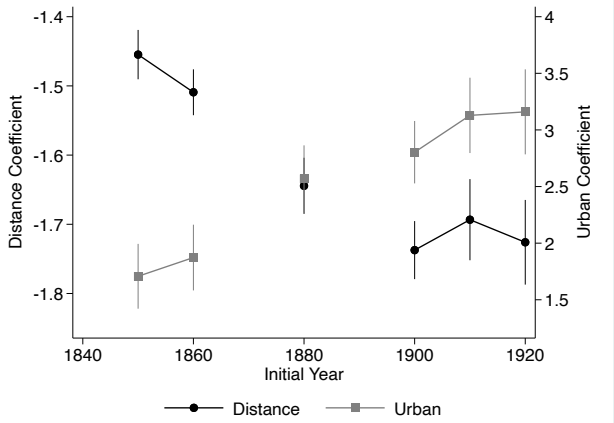
Figure 7: Changes in distance of moves

Note: These figures show the coefficients on initial-year fixed effects in regressions of destination characteristics with and without controls, with 1850 as the excluded year. "Controls" includes all controls available in all censuses, including state fixed effects. "Controls with FE" includes also county fixed effects. Vertical lines are 95-percent confidence intervals. Observations weighted to correct for selection into linkage.

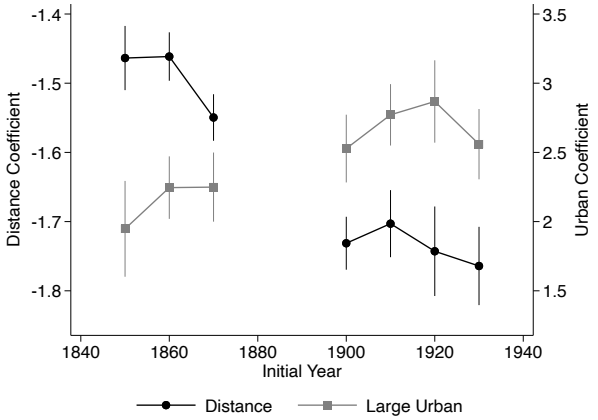
(a) Urban (2,500+), 10-year



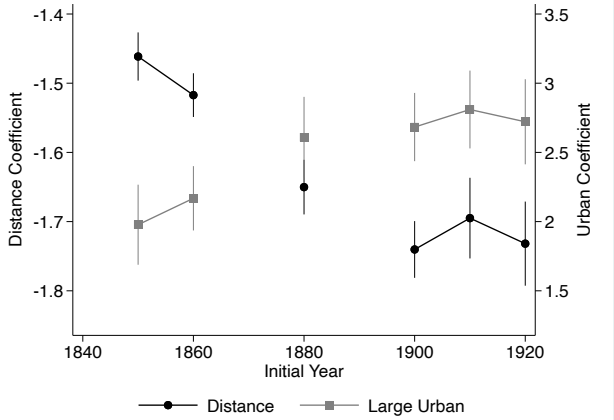
(b) Urban (2,500+), 20-year



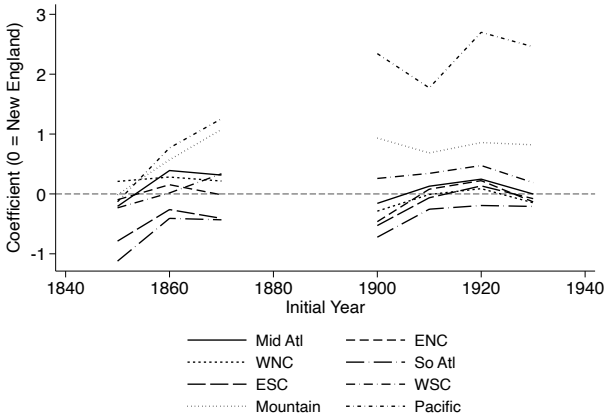
(c) Urban (25,000+), 10-year



(d) Urban (25,000+), 20-year



(e) Census divisions, 10-year



(f) Census divisions, 20-year

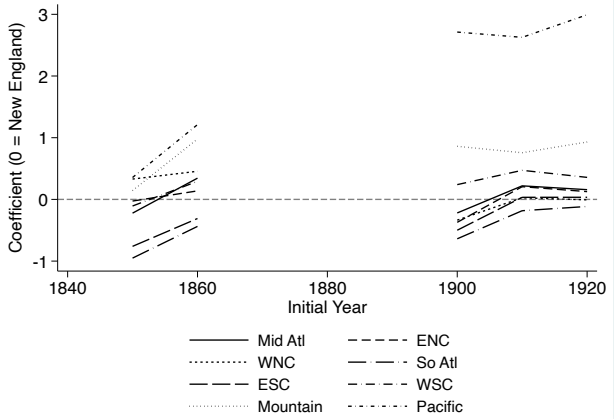
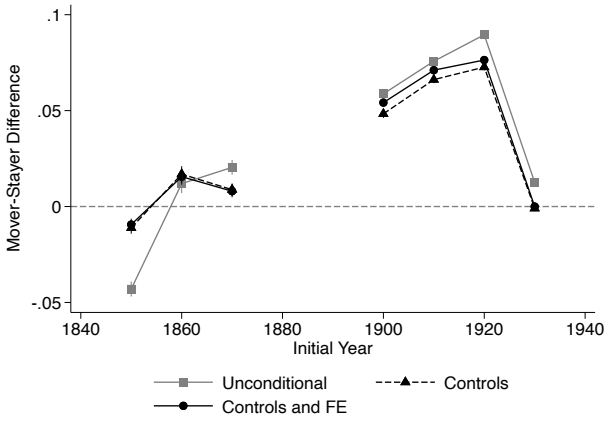


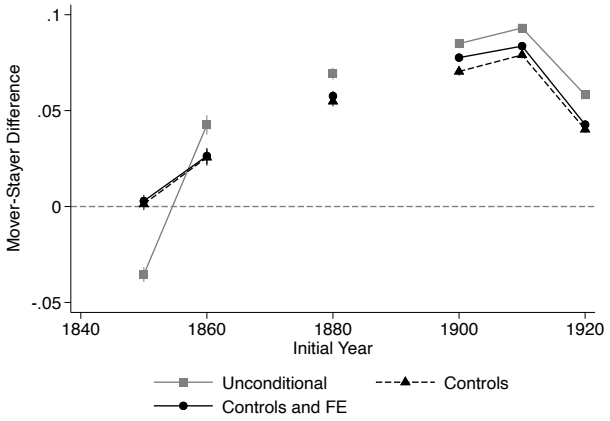
Figure 8: Conditional logit results

Note: Panels (a)–(d) present coefficients on distance and the measure of urban residence from the conditional logit estimation. Panels (e) and (f) present the coefficients from the regressions of panels (a) and (b), respectively, on the census division indicators, with New England excluded.

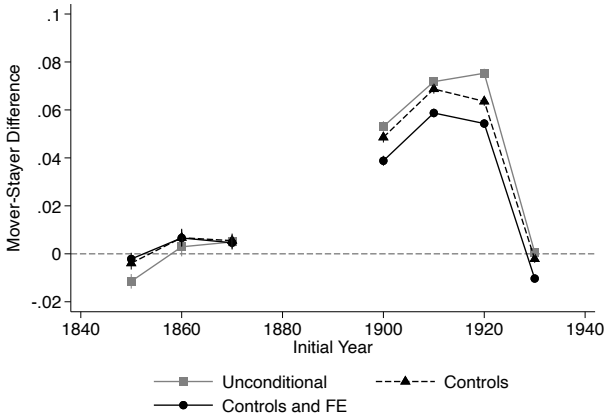
(a) Urban (2,500+), 10-Year



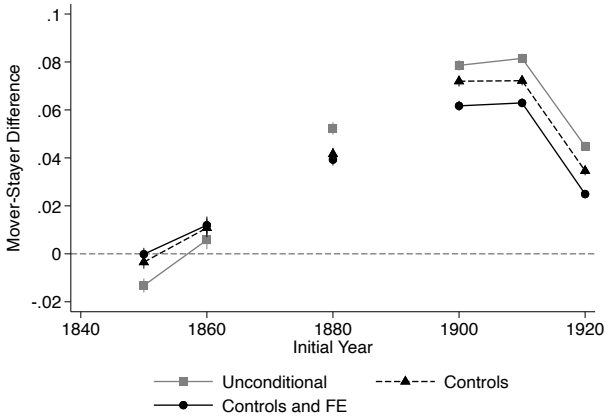
(b) Urban (2,500+), 20-Year



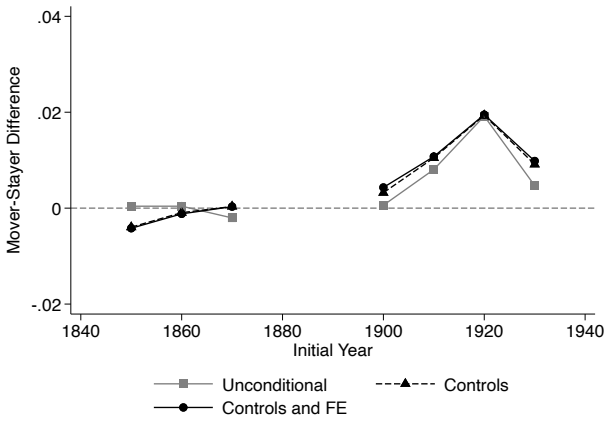
(c) Urban (25,000+), 10-Year



(d) Urban (25,000+), 20-Year



(e) Bartik, 10-Year



(f) Bartik, 20-Year

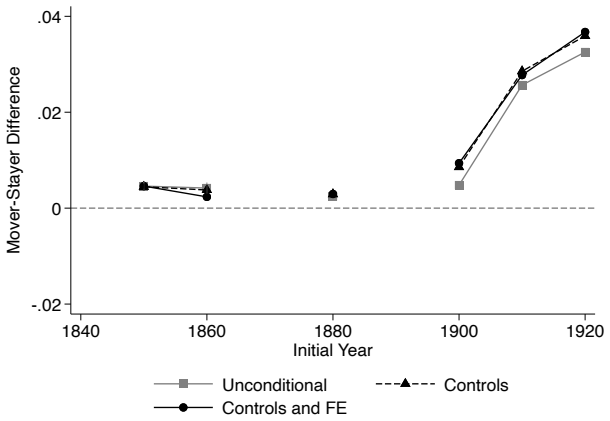


Figure 9: Movers' change in urbanization and Bartik

Note: This figure compares the growth in urbanization or labor demand experienced by movers to that experienced by stayers, with either no controls, controlling for all observables, and controlling for all observables and county fixed effects.