

NBER WORKING PAPER SERIES

EUROPE'S TIRED, POOR, HUDDLED MASSES:
SELF-SELECTION AND ECONOMIC OUTCOMES IN THE AGE OF MASS MIGRATION

Ran Abramitzky
Leah Platt Boustan
Katherine Eriksson

Working Paper 15684
<http://www.nber.org/papers/w15684>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2010

We have benefited from conversations with Timothy Bresnahan, Moshe Buchinsky, Dora Costa, Pascaline Dupas, Joseph Ferrie, Claudia Goldin, Avner Greif, Timothy Guinnane, Rick Hornbeck, Seema Jayachandran, Lawrence Katz, Naomi Lamoreaux, Shirlee Lichtman, Robert Margo, Roy Mill, Joel Mokyr, Paul Rhode, Kjell Salvanes, Izi Sin, Gunnar Thorvaldsen, Gui Woolston, Gavin Wright and members of the KALER group at UCLA. We thank seminar participants at Harvard, Humboldt, Queen's, Simon Fraser, Toronto, Warwick and Yale, as well as conference participants at the Economic History Association, the Nordic Labor Economics Summer Institute, the Social Science History Association and the Development of the American Economy and Labor Studies groups at the NBER. Matthew Baird and Roy Mill helped to collect data from Ancestry.com. John Parman and Sula Sarkar generously shared data with us. We acknowledge financial support from the National Science Foundation, the California Center for Population Research and UCLA's Center for Economic History. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2010 by Ran Abramitzky, Leah Platt Boustan, and Katherine Eriksson. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Europe's tired, poor, huddled masses: Self-selection and economic outcomes in the age of mass migration

Ran Abramitzky, Leah Platt Boustan, and Katherine Eriksson

NBER Working Paper No. 15684

January 2010. "Tgxkgf 'P qxgo dgt"4232

JEL No. J61,N31

ABSTRACT

The Age of Mass Migration (1850-1913) was among the largest migration episodes in history. During this period, the United States maintained open borders. Using a novel dataset of Norway-to-US migrants, we estimate the return to migration while accounting for migrant selection across households by comparing migrants with their brothers who stayed in Norway. We also compare the fathers of migrants and non-migrants by wealth and occupation, and examine migrants' assimilation in the US labor market. We find that, unhindered by entry restrictions, migrants were negatively selected from the sending population and their return to migration was relatively low.

Ran Abramitzky
Stanford University
Economics department
579 Serra Mall
Stanford, CA 94305
ranabr@stanford.edu

Katherine Eriksson
Department of Economics
8283 Bunche Hall
UCLA
Los Angeles, CA 90095-1477
kath722@ucla.edu

Leah Platt Boustan
Department of Economics
8283 Bunche Hall
UCLA
Los Angeles, CA 90095-1477
and NBER
lboustan@econ.ucla.edu

“Keep, ancient lands, your storied pomp!” cries she
With silent lips. “Give me your tired, your poor,
Your huddled masses yearning to breathe free,
The wretched refuse of your teeming shore.
Send these, the homeless, tempest-tost to me,
I lift my lamp beside the golden door!”

- Emma Lazarus, “The New Colossus” (1883)
Displayed upon the Statue of Liberty in New York Harbor

1. Introduction

The Age of Mass Migration from Europe to the New World was one of the largest migration episodes in human history. Between 1850 and 1913, the United States absorbed nearly 30 million European immigrants. This paper asks two related questions about this migrant flow. What was the economic return to migrating from Europe to the United States in the late nineteenth century? And, were migrants positively or negatively selected from the European population? We study whether the US acquired wealthier and higher-skilled European migrants who were able to finance the voyage or whether it absorbed Europe’s “tired, poor, huddled masses” who migrated to the US in search of opportunity.

Understanding migration in this era is of particular importance. First, the US maintained a nearly open border in the late nineteenth century, allowing us to study migrant self-selection and the economic return to migration without interference from the bureaucratic factors that govern migrant selection today. In contrast, in the current period, the immigrant flow is a product not only of individual migration decisions, but also of complicated entry rules and restrictions, which obscure the underlying economic forces. Thus, comparing our findings with contemporary studies can illuminate the effect of modern immigration policy on migrant selection. Second,

given the large magnitude of the migration flow, the skill composition of departing migrants had potentially large implications for economic growth in Europe and the US.

Our empirical methods are also of interest to labor economics and the economics of migration. Because migrants are not randomly selected from the sending population, it is challenging to separately identify the return to migration and the selection into migration. Measuring the return to migration with a naïve comparison of migrants and stayers can be confounded by migrant selection. For example, migrants may have earned more than men who remained in Europe in part because the brightest people, who would have earned more regardless of location, were the most likely to move to the US. Therefore, in the presence of positive selection, a naïve OLS estimate of the return to migration will be biased upward, and, similarly, in the presence of negative selection it will be biased downward.

To account for across-household selection into migration, we compare the earnings of migrants to the earnings of their brother(s) who remained in Europe. The resulting estimate of the return to migration eliminates the across-household component of migrant selection, which can result from differing propensities to migrate from households that are financially constrained or face poor economic opportunities in Europe. Furthermore, this within-household estimate eliminates the component of unobserved ability that is shared between brothers.

Beyond providing a more accurate estimate of the return to migration, this method allows us to infer the nature and extent of migrant selection across households. Specifically, a comparison of the within-household estimate and the naïve OLS estimate reveals the direction of the across-household component of migrant selection. For example, if the naïve OLS estimate of the return to migration were smaller than the within-household estimate, this contrast would

suggest that the naïve OLS estimate was biased downward by negative selection of migrant households.

We focus on Norwegian migrants to the US for two reasons. First, Norway had one of the highest out-migration rates among European sending countries, with over a quarter of its population eventually migrating to the US. Second, Norway has completely digitized two censuses from the period (1865 and 1900), allowing us to follow large samples of men over time. We create a novel data set of all Norwegian-born men in the US in 1900 using US Census records from the genealogy website, Ancestry.com. We then match men by name and age from their birth families in Norway in 1865 to the labor market in either Norway or the US in 1900.

The outcome we observe for each individual is an occupation either in the US or in the Norwegian labor market.¹ We then assign individuals the mean earnings for their occupation in either Norway or the US (in real PPP-adjusted 1900 US dollars). For simplicity, we often refer to this occupation-based earnings measure as “earnings,” but it should be thought of as an occupational ranking. While this measure captures two components of the return to migration, namely the potential for higher mean earnings in the US for each occupation and the potential for occupational upgrading, it cannot account for the potential for a higher return to skill *within* occupation in the US.² Despite this drawback, the historical data have an important advantage over their modern counterparts. Due to privacy restrictions, the individual names that we use to match migrants to their birth families are only released 70 or more years after the initial Census

¹ In principle, one could also study migrant selection by comparing the education levels or literacy rates of migrants to men who remained in Norway. However, the Norwegian Census did not collect information on literacy or years of schooling in 1900. 97 percent of Norwegian-born men in the relevant age range who are observed in the US Census in 1900 report being literate.

² To the best of our knowledge, there are no data sources that would allow us to measure variation in earnings within an occupation in either the US or Norway circa 1900.

was taken, rendering historical Census data the only large data set available for sibling comparisons of migrants and non-migrants.

We estimate a return to migration of around 60 percent after accounting for the selection across households into migration.³ Such returns are lower than contemporary estimates for the return to migration from Mexico to the United States (200-400 percent; see Hanson, 2006). We also find evidence of negative selection, suggesting that men with poorer economic prospects in Norway were more likely to move to the US in the late nineteenth century.⁴ The difference between the within-household estimates and the naïve OLS estimate suggests that negative selection leads the naïve OLS estimate of the return to migration to understate the true return by 20 to 50 percent for migrants leaving urban areas.

We gather additional information to study the nature of migrant selection. First, we compare the occupation-based earnings, asset holdings and property tax revenues of the fathers of migrants and non-migrants in Norway. We find that the fathers of migrants are poorer than those of non-migrants. Second, we examine the assimilation patterns of Norwegian migrants in the US labor market, and observe that Norwegian migrants earn lower wages than natives even after many years in the US. These patterns provide additional suggestive evidence for negative selection in migration from both rural and urban areas.

Our finding that migrants were negatively selected from the sending population is consistent with Borjas' (1987, 1991) work on migrant selection in the Roy model. Borjas predicts negative selection when migrants move from origins with more unequal income

³ We capture the return to migration at a specific point in the US-to-Norway migration. Ultimately, we expect wages in the two countries to converge as out-migration reduces the labor supply in the sending country (O'Rourke and Williamson, 1995, 1999, 2004). Therefore, the return to migration would likely fall over time as the two countries experience this convergence process.

⁴ Again, we note that we can only measure selection across occupations. We acknowledge that migrants may have been the brightest and most motivated men holding these low-ranked occupations.

distribution to destinations with less unequal income distribution. Unlike today, Norway had a more unequal income distribution in the nineteenth century than did the US (see Figure 1).

The remainder of the paper proceeds as follows. Section 2 discusses the historical context and related literature on the Age of Mass Migration and migrant selection. Section 3 describes the data and the procedures we used to match migrants to their birth families in Norway. Section 4 presents our estimates of the return to migration, while Section 5 contains additional direct evidence of migration selection. Section 6 concludes.

2. Contemporary and historical literature on migrant selection

A. Migrant selection in the Roy model

The Borjas model of migrant selection is both well-known and much-disputed in the migration literature. Borjas (1987, 1991) modified the Roy model of occupational choice to generate predictions about the nature of migrant selection (Roy, 1951).⁵ In this framework, migrant selection is determined by the relative return to skill in the sending and destination economies. If the destination country exhibits higher return to skill than the source country, and therefore greater levels of income inequality, migrants will be disproportionately drawn from the top end of the source country's skill distribution. If, instead, the destination country offers lower return to skill and is therefore more equal than the source, migrants will be disproportionately drawn from the lower tail of the source country's skill distribution.

Work on contemporary immigrant flows has found only mixed support for the Borjas model of migrant selection. Gould and Moav (2010) show that Israeli migrants to the United States are positively selected, at least on observable skills, as would be predicted by the more

⁵ For an alternative view on migrant selectivity, see Chiswick (1999, 2000).

compressed distribution of Israeli wages.⁶ However, Chiquiar and Hanson (2006) observe that Mexican migrants to the United States are drawn from the middle, rather than the low end, of the Mexican skill distribution, despite the fact that income inequality is higher in Mexico than in the United States.⁷ Moreover, Feliciano (2005) and Grogger and Hanson (2008) find that migrants are positively selected on educational attainment from almost every sending country in the world, even those countries with very high levels of income inequality.

Scholars have attempted to reconcile the Borjas model with the facts about positive selection in a variety of ways. A new generation of models incorporates borrowing constraints and shows that, as the cost of migration increases, the poorest residents of sending countries can no longer afford to move (Borger, 2009; McKenzie and Rapoport, forthcoming).⁸ Alternatively, Grogger and Hanson (2008) demonstrate that a classic Roy model with a linear, rather than a logarithmic, utility function generates predictions of positive selection whenever the skill-related differences in *wage levels*, rather than the relative return to skill *in percentage terms*, are high. In this framework, positive selection is a likely outcome in the contemporary world given the dramatic difference in wage levels between developed and developing countries.

B. The Age of Mass Migration

Between 1850 and 1913, more than 40 million Europeans moved to the New World, nearly two-thirds of whom settled in the United States.⁹ Initially, migrants from the British Isles and Germany constituted the majority of the migrant flow to the US. These early migrants were

⁶ See also Abramitzky (2008) and Borjas (2008), which find support for the Borjas selection hypothesis in the contexts of migration to and from Israeli *kibbutzim* and Puerto Rico, respectively.

⁷ See also a series of papers on migration from the Pacific Islands (Akee, 2010; McKenzie, Gibson, and Stillman, forthcoming; McKenzie and Gibson, forthcoming).

⁸ The cost of migrating from Mexico to the US is around \$2000 in 2000 US dollars, or 35 percent of the annual earnings of a low-skilled Mexican worker (Hanson, 2006).

⁹ This paragraph is based on Hatton and Williamson (1994, 1998).

joined by Scandinavians and other Northern Europeans in the 1870s and by Southern and Eastern Europeans in the 1880s. Norway experienced one of the highest out-migration rates in the 1880s, during which time 10 of every thousand Norwegians left the country on an annual basis.

With the shift from sail to steam technology on the Atlantic, the cost of migration fell dramatically over the nineteenth century (Keeling, 1999). The declining cost of migration, coupled with rising real incomes in the newly industrializing European periphery, relaxed the financial constraints on households that previously had been too poor to pay for passage to the New World. As migration became affordable to a greater share of the European population, the migrant flow shifted from the richer countries of Western Europe to the poorer countries in the South and East (Hatton and Williamson, 1998, 2006; O'Rourke and Williamson, 1999, 2004).

Beyond these broad “macro” patterns, we know very little about the characteristics of individuals who selected to leave Europe and move to the New World in the 19th century. For example, within a country, was the migrant flow drawn from the top or bottom end of the skill distribution? What was the economic return to this migration, after accounting for migrant self-selection? To our knowledge, Wegge (1999, 2002, 2010) are the only papers to provide individual-level evidence on migrant selection in the nineteenth century.¹⁰ Wegge documents intermediate selection for the emigration flow leaving Germany in the 1850s: members of the highest- and lowest-skill occupations were less likely to migrate than were workers in the mid-skill range, such as machinists, metal workers and brewers. She concludes that the poorest migrants may have lacked the resources necessary to finance their trip. This result may be specific to the 1850s when the cost of passage to the New World exceeded the total annual earnings of the average German laborer. Following Hatton, O'Rourke and Williamson's logic,

¹⁰ For work on migrant selection in other historical periods, see Ferrie (1999) on rural-to-urban migration in the US, Margo (1990) on black migrants leaving the US South, and Abramitzky and Braggion (2006) on indentured servants to New World colonies.

we would expect the later Norwegian migration to be more negatively selected than the earlier German flow.

An application of the Roy model to our historical context also generates a prediction of negative selection. In the modern context, the Scandinavian countries are more equal than the United States. However, in the late nineteenth and early twentieth century, the opposite was true. Figure 1 compares the occupation-based earnings distribution in the United States and Norway in 1900. We array individuals from lowest- to highest-paid with earnings represented in US dollars and the Norwegian distribution rescaled to share the US mean (the earnings data are described in more detail in Section 3). US workers below the 50th percentile of the earnings distribution out-earned similar Norwegians, while Norwegians above the 90th percentile commanded higher earnings than their US counterparts. These occupation-based earnings distributions suggest that Norway offered a higher return to skill than did the United States circa 1900, which is consistent with the historical evidence on 90-50 ratios in the two countries (Soltow, 1965; Goldin and Katz, 1999).¹¹

Not only was the occupation-based earnings distribution in the US more compressed at a point in time but the US economy also offered the opportunity for substantial occupational upgrading over the life-cycle. Ferrie and Long (forthcoming) document that only 18 percent of men in the US who held an unskilled, blue collar job in 1850 remained unskilled workers by 1880. By comparison, 47 percent of men in unskilled, blue collar occupations in Norway in 1875

¹¹ Soltow (1965) compares the average earnings for men in the top decile of the income distribution to mean earnings in urban places in Norway in 1890. He finds a [90-100]/mean ratio of 4.25. Goldin and Katz (1999) instead calculate a more conventional 90/50 ratio of 1.71 for 12 urban industries in the US in 1890. To compare these two figures, we use two adjustment factors: (1) the ratio of median to mean income of 0.93 from the 1911 Canadian Census (Green and Green, 2008) and (2) the ratio of average earnings in the top decile to earnings at the 90th percentile of 2.56 from the 1917 US tax returns (Piketty and Saez, 2003). By this method, the pseudo [90-100]/mean ratio for the US in 1890 is 4.06, which is a bit lower than Norway. We note that our adjustment factors may inflate the US ratio, particularly because top-end inequality in the US was likely higher in 1917 than in 1890. Even by this conservative measure, we find that Norway was less equal than the US.

remained unskilled workers in 1900.¹² Men who start their careers in unskilled occupations were twice as likely to move up the occupational ladder in the US than in Norway over their lifetime; much of this mobility was accomplished by moving into owner-occupier farming.

In historical terms, the costs of migration were relatively low in the late nineteenth century. We estimate that the total cost of migration, including foregone earnings during the voyage, represents around 18 percent of the annual earnings of a Norwegian farm laborer.¹³ Migrant networks also helped to defray the cost of passage for new arrivals; 40 percent of Norwegian migrants during this period travelled on pre-paid steamship tickets financed by friends or relatives (Hvidt, 1975, p. 129).

3. Data and matching

A. Matching Norwegian-born men to their birth families

Our goal is to create a dataset of Norwegian migrants and non-migrants whom we can observe at two points in time, first when living in their childhood household and then when participating in the labor market later in life. We rely on three Census sources: the complete digitized Norwegian Censuses of 1865 and 1900 and a novel data set containing the full population of Norwegian-born men in the US in 1900. The Norwegian Census data are archived by the North Atlantic Population Project (NAPP). We compiled a complete roster of Norwegian immigrants living in the US in 1900 from the genealogy website Ancestry.com, which contains some digitized variables from historical Census manuscripts.

¹² We derive this result by matching men between the 1875 and 1900 Norwegian Censuses using the matching procedure outlined in section 3a.

¹³ Norwegian farm laborers earned around \$175 in 1900 US dollars. For this calculation, we assume that migrants lost 20 days of work for the passage and the resettlement. However, it is interesting to note that Armstrong and Lewis (2009) report that the typical Dutch migrant to Canada in the 1920s saved around \$150 (in 1900 US dollars) for the cost of the voyage and resettlement, nearly a full year's salary for a Norwegian farm laborer.

We match men young enough to live in their childhood household in the 1865 Norwegian Census to the population of Norwegian-born men observed in Norway or the US in 1900. The Norwegian and US Censuses of 1900 are combined to create the universe of Norwegian-born men in that year. Because over 95 percent of emigrants from Norway settled in the United States, these two sources contain nearly all Norwegian-born men who survived to 1900 (Ferenczi and Willcox, 1929). Our baseline method (“**Match 1**”) uses an iterative matching strategy pioneered by Ferrie (1996).

We describe this procedure in detail:

- (1) We identify 257,767 Norwegian men between the ages of 3 and 15 in 1865. 72,798 of these men are unique by first name, last name and birth year in 1865.
- (2) We standardize all first and last names in both datasets to address orthographic differences between phonetically equivalent names using the NYSIIS algorithm (see Atack and Bateman, 1992).
- (3) We match unique observations in 1865 forward to 1900 using an iterative procedure. We start by looking for a match by name and exact birth year. If we find a *unique* match here, we stop and consider the observation “matched.” If we find multiple matches for the same birth year, the observation is thrown out. If we do not find a match at this first step, we try matching first within a one-year band (older and younger) and then with a two-year band around the reported birth year. If neither of these attempts produces a match, the observation is considered to be “unmatched.”¹⁴

This procedure generates a sample of 2,826 migrants and 18,820 non-migrants. We achieve a forward match rate of 30 percent, which is comparable to Ferrie and Long’s

¹⁴ We restrict our attention to men who are at least three years old in 1865 to ensure that all observations can match to a two-year age band around the reported age.

(forthcoming) forward match rate of 22 percent within the United States over a similar 30 year period (1850-80).¹⁵ Some matches fail because of name-age combinations that cannot be found in 1900 and others because of name-age combinations that are not unique in 1900.¹⁶

We can expand our sample size further by adding province of birth as a third match criterion for men who remain in Norway in 1900. Matching by name, age and province of birth allows us to differentiate between men who would otherwise not be considered unique, thereby increasing our sample size to 40,357. Although this approach (“**Match 2**”) has the advantage of a larger sample size, it may introduce a bias by using different matching procedures for migrants and stayers. For this reason, Match 1 is our preferred approach.

We note that the iterative nature of this method may produce false matches because some men may have both an exact match and a “close” match (within a one- or two-year band around the reported birth year). As a robustness exercise, we match a restricted sample of men who are unique by name within a *five-year age band* in both Censuses (two years around the reported age in each direction). This approach (“**Match 3**”) limits the potential for false matches in 1900 but also reduces the sample size to 9,423.

¹⁵ If we instead conducted a backwards match from 1900 to 1865, we achieve a match rate of 23 percent, a weighted average of 11 percent for men living in the US in 1900 and 25 percent for men living in Norway. The differential match rate between migrants and non-migrants could be due to the practice of anglicizing one’s name upon arrival in the US, a factor that we consider in the next section.

¹⁶ Of the 72,798 potential matches in 1865, we are able to match 30 percent in 1900, fail to match 4 percent due to name-age combinations that are not unique in 1900 and fail to match the remaining 66 percent due to name-age combinations that cannot be found in 1900. Some of the match failures due to name-age doubles in 1900 reflect age misreporting. Indeed, when we expand our age band to 10 years, we locate an additional 4,328 individuals, suggesting that 6 percent of potential matches misreported their age by a few years ($=4,328/72,798$). All of our results are qualitatively similar if we expand our age band to 10 years. Match failures due to missing name-age combinations could be due either to mortality between 1865 and 1900 or to name changes and transcription error. We use US mortality rates by age and sex in 1900 to calculate an expected 35-year mortality rate of 25 percent. We conclude that mortality can account for around 40 percent of the match failures due to missing name-age combinations.

B. Occupation and earnings data in Norway and the United States

We observe labor market outcomes in 1900, when the men in our sample are in their 30s and 40s. Neither the US nor the Norwegian Census of 1900 contains individual information on wages or income. Instead, we assign men the mean income earned by members of their occupation.¹⁷ Men living in the United States are matched to income data from the 1901 Cost of Living Survey while men living in Norway are matched to mean income-by-occupation tabulations for the year 1900 published by Statistics Norway and other sources (Haines and Preston, 1991; *Statistik Aarbog*, 1900; Grytten, 2007).¹⁸ The 1901 Cost of Living Survey reports income information for more than 300 occupations in the US. Our dataset contains individuals representing one hundred and eighty-nine occupational categories. We convert Norwegian wages to real, PPP-adjusted US dollars using the 1900 exchange rate and price levels reported in Grytten (2004). Appendix A provides more detail on the data sources and assumptions underlying these estimates.

The 1901 Cost of Living survey may overstate the return to migration both because the survey was conducted in urban areas and because the majority of survey respondents were native born. We address these concerns by considering alternative sources of earnings data in the US. First, we calculate earnings by occupation from the 1915 Iowa Census, which better represents the urban/rural divide in our US sample. In addition, we calculate the wage gap between Scandinavian migrants and native-born workers in the US circa 1900 using data from the Immigration Commission and the Census. We estimate that Scandinavian migrants earned 13 log

¹⁷ For men living in the US, we code occupation by hand using the digital images of Census manuscripts available on Ancestry.com.

¹⁸ *Statistics Norway* reports daily wage rates. We convert these wage rates into annual earnings figures by assuming that Norwegians worked six-day work-weeks and were unemployed for 0.66 months during the year (= 297 days of work per year, on average). Our estimate for months spent unemployed is based on reported unemployment for Norwegian migrants in the 1900 US Census.

points less than native workers in their occupation.¹⁹ In some specifications, we reduce the earnings assigned to Norwegian migrants in the US by this amount.²⁰

Table 1 reports the ten most common occupations for our sample of matched migrants in the United States and matched stayers in Norway. Although forty percent of both groups report working in farm occupations, migrants to the US were much more likely to be owner-occupier farmers (36 percent versus 22 percent). Migrants were also more likely to report being general laborers (8 percent versus 1.4 percent). Other common occupations in both countries include carpenters, fishermen and sailors.

Our unavoidable reliance on mean earnings by occupation prevents us from measuring the full return to migration. Conceptually, the return to migration derives from three channels: (1) the presence of higher wages in the US in the typical occupation; (2) the possibility that migrants are able to switch from low-paying to high-paying occupations upon arriving in the US; and (3) the potential for higher within-occupation return to ability in the US. Our estimate of the return to migration cannot capture the third aspect of the total return because we do not observe variation in earnings within an occupation.

We face a related limitation in our ability to describe the extent of migrant selection. Positive selection, for instance, could be generated either by high migration rates among men

¹⁹ According to the Immigration Commission, Scandinavian migrants earned 15 log points below native-born workers in the same industry (Hatton and Williamson, 1998, pp. 146-7). This wage penalty reflects not only the fact that, within industries, migrants may have held lower-paying occupations but also that migrants may have earned less than natives even within a given occupation. Using supplemental Census data, we infer that the majority of this earnings penalty (13 log points) was due to within-occupation differences in wages. In particular, we use the 1900 IPUSM sample to run a regression of our (log) occupation-based earnings measure on being born in Scandinavia and industry fixed effects for the 16 narrowly-defined mining and manufacturing industries reported in the Immigration Commission data. The Scandinavia coefficient is -0.018 (p-value = 0.102), leading us to conclude that all but 2 log points of the 15 point wage penalty appears to have been due to within-occupation differences in wages.

²⁰ However, some portion of the 13 log point wage gap could be due to the fact that migrants are negatively selected. That is, perhaps migrants' earnings would have been in the lower tail of the wage distribution in their occupation regardless of whether they lived in Norway or the US. In this case, we would not want to adjust the return to migration for (all of) this 13 log point wage gap. As a result, we choose not to highlight this specification as our preferred estimate of the return to migration.

from occupations with high mean earnings or by high migration rates among men at the 80th or 90th percentile of the wage distribution *within* their occupation. The reverse is true, of course, for negative selection. With our data, we can document the fact that more (fewer) common laborers moved to the US, but we will not be able to observe whether the best (worst) among the laborers made the journey.

C. The occupational distribution of migrants vs. stayers

This section compares the occupational distributions of Norwegian migrants to the US and men who remained in Norway in 1900. Figure 2 presents the occupational distributions of these migrants and stayers, with occupations arrayed from lowest- to highest-paid according to the average US earnings in that occupation.²¹ We omit farmers, the largest occupational category, for reasons of scale but results are qualitatively similar when farmers are included. For men born in urban areas, migrants are more likely to hold low-paying jobs such as day laborer or servant, while the men remaining in Norway exhibit an occupation distribution that is skewed toward higher-paying jobs (for example, merchants). Men born in rural areas are employed in similar jobs in both countries.

After constraining occupations to have the same mean earnings in Norway and the US, we find that, on average, urban migrants sorted into occupations paying 19 log points less than the typical urban dweller in Norway. This negative “return to migration” is consistent with either initial negative selection or occupational downgrading in the US or both.²² Disadvantages

²¹ Chiquiar and Hanson (2006) conduct a similar exercise for Mexican migrants to the United States using the 2000 Census. They assign migrants the earnings that they would have received, given their education and experience level, if they had remained in Mexico. We use US earnings, rather than Norwegian earnings, because the US earnings data are richer, reflecting nearly 200 occupational categories.

²² Norwegian migrants may have experienced occupational upgrading or downgrading in the US for various reasons. On the one hand, higher rates of occupational mobility in the US may have allowed migrants to climb the occupational ladder. In particular, given the low price of land in the US, many workers who started out as

resulting from a lack of US-specific skills would likely erode over time as migrants learn English and invest in other country-specific skills. However, we find a similarly negative “return to migration” (18 log points) for urban migrants who had been in the US for 20 years or more. Therefore, we suspect that occupational differences between migrants and stayers are unlikely to be driven by temporary occupational downgrading.

D. Comparing matched samples with the full population

Our matched samples may not be fully representative of the population of Norwegian-born men, either in 1865 or in 1900. In particular, matched samples are selected for having uncommon first and last names, which may have been associated with higher socio-economic status. Table 2a compares the attributes of men in the primary matched sample to the Norwegian population in the same age range in the 1865 Census, while Table 2b compares matched migrants to Norwegian-born men living in the United States and matched stayers to men living in Norway in 1900.²³

By construction, men with uncommon names are more likely to be successfully linked between Censuses. Table 2a shows that the median rural man in the population shared his first and last name with 121 others, while the median urban man shared his name with only 11 others. Unsurprisingly, name frequency in the matched sample is substantially lower than that of the population, with rural men sharing their name with 7 others and urban men with only 2 others in the Norwegian population.

agricultural laborers were able to purchase their own farm and become an owner-occupier. On the other hand, though, migrants may have lacked the US-specific skills necessary to secure a highly-paid occupation.

²³ Appendix A (section D) compares the matched samples to subgroups of the population by cause of match failure. For these tables and the remainder of these analyses, we drop men who lived in group quarters in 1865 (1,676 men in our matched sample) because we are unable to reconstruct aspects of their childhood households.

Our matched sample is demographically similar to the population in terms of age, number of siblings, and birth order. However, men in our matched sample are more likely to live in urban areas, both in childhood and as adults, and to hail from households of somewhat higher socio-economic status than the population average.²⁴ According to Table 2a, only 14 percent of Norwegian men lived in an urban area in 1865 compared with 26 percent of our matched sample. Within urban areas, matched men are 10 percentage points more likely to live in a household whose head holds an occupation with above-median earnings and twice as likely to live in a household with some assets, defined as owning a business or serving as a master craftsman in an artisanal workshop. In rural areas, matched men are also drawn from higher socio-economic status households although this difference is not as pronounced. By adulthood, this privilege translates into slightly higher labor market earnings: men in our matched samples earn around 4 percent more than the comparable population in 1900, both among migrants to the US and men who remain in Norway (Table 2b). As we expected, having an uncommon name appears to be a proxy for urban location and socio-economic status, perhaps because urban households used a wider array of given names (Gjerde, 1985, p. 48).

Although our matched sample is not fully representative of the population from which it is drawn in terms of place of residence (urban/rural) and socioeconomic status, three things are worth noting. First, the direction and extent of this bias is nearly identical both for migrants and stayers. Table 2b shows that men in our matched sample earn around 4 percent more than the average Norwegian-born man both in Norway and in the US in 1900. Therefore, the distinctive features of our matched sample are not likely to affect our estimates of the economic return to migration or our conclusions about migrant selection, which depend on a comparison of matched

²⁴ Norwegian households were defined by the Census as urban if their municipality of residence was considered to be a town. However, many towns contained agricultural land on their periphery. Therefore, the urban designation likely includes some households with “rural” characteristics.

migrants to matched stayers rather than a comparison of the matched sample to the population. Second, throughout the paper we split our sample into urban and rural and conduct our analysis separately for men hailing from urban and rural areas. Finally, to further reduce the differences between our matched sample and the general population, we consider specifications that reweight our matched sample to resemble the frequencies of the following characteristics in the general population: urban residence, asset holdings and above-median occupation of the household head.

We address a few additional limitations of our matching procedure here. First, our method will not capture migrants who anglicize their name upon arrival in the US, which could be a concern if changing one’s name is correlated with economic success. Following Fryer and Levitt (2004), we use the complete 1880 US Census to construct indices of a name’s distinctively Norwegian content.²⁵ By this metric, we find that men in our matched sample are no more likely than the typical migrant to have a distinctively Norwegian name (see Table 2b, row 7). This pattern likely reflects a tradeoff: although we fail to match men who adopt less Norwegian names in the US, our matching procedure initially selects for men who have uncommon names (that is, names that are rare in Norway). In addition, we find no evidence that the “Norwegian-ness” of a man’s name is related to our occupation-based earnings measure.²⁶

Second, our sample of matched migrants will not include temporary movers who returned to Norway before 1900. According to the aggregate statistics, 25 percent of the Norwegian

²⁵ Our name index ranges from zero to two, with a value of zero reflecting the fact that no men in the US with a given first and last name were born in Norway and a value of two assigned to men whose first and last names are both distinctively Norwegian. The first name index is equal to

$$\frac{\text{Pr}[1^{\text{st}} \text{ name} | \text{Norwegian_born}]}{\text{Pr}[1^{\text{st}} \text{ name} | \text{Norw_born}] + \text{Pr}[1^{\text{st}} \text{ name} | \text{not_Norw_born}]}$$

and likewise for the last name index. The full measure adds these two indices together.

²⁶ We regress $\ln(\text{earnings})$ on the full name index and a quadratic for age for Norwegian-born men in the 1900 IPUMS in the relevant age range. The coefficient on the name index is 0.018 (s.e. = 0.017). By this estimate, the average difference in the index value of 0.05 between matched and unmatched men would translate into a 0.1 percent difference in earnings, which is both small and statistically insignificant.

migration flow eventually returned to Norway (Semmingesen, 1978, p. 20).²⁷ Return migrants may have been disproportionately drawn from the upper or lower end of the income distribution, either because unsuccessful migrants return home to lean on their familial support network or because the most successful migrants are able to build up a certain level of savings most quickly in order to return home. The availability of an intermediate US Census in 1880 allows us to address this point. We identify over 25,000 Norwegian-born men in the relevant age range in the 1880 Census. We are able to locate 14 percent of these men in either the US or the Norwegian Censuses of 1900; one-third of these had returned to Norway. We compare the economic outcomes of migrants who eventually returned to Norway and those who remained in the US in 1880, when both sets of migrants were still living in the US. Figure 3 reveals few discernable differences in the occupational distributions of these two groups.²⁸ Men who eventually returned to Norway are slightly over-represented at the bottom end of the occupational distribution but the mean occupation scores of returners and persisters are statistically indistinguishable.

4. Estimating the return to migration in the presence of selection

A. A naïve estimate of the return to migration: The mean earnings of migrants vs. stayers

One naïve approach to estimating the return to migration is to compare the occupation-based earnings of all Norwegian-born men living in the United States to all men in Norway in 1900:

$$\ln(\text{Earnings}_i) = \alpha + \beta_1(\text{Migrant}_i) + \beta_2(\text{Age}_i) + \beta_3(\text{Age}_i^2) + \varepsilon_i \quad (1)$$

²⁷ The United States only began tracking return migration in 1907-1908. Gould (1980) reports a much lower return migration rate (6.7 percent) for Norwegians for the 1907-1913 period.

²⁸ We compare these groups using the occupation score variable available in the 1880 IPUMS data, rather than our occupation-based earnings measure. The occupation score variable is constructed in a similar manner by matching occupations to their median earnings in 1950.

where $Earnings_i$ denotes the mean earnings of members of individual i 's occupation in 1900 in his country of residence, $Migrant_i$ is a dummy variable equal to one if individual i lives in the United States in 1900, and Age_i and Age_i^2 are individual i 's age and age-squared in 1900.²⁹ The US Census data are taken from the Integrated Public-Use Microdata Series (IPUMS).³⁰ For now, we measure the return to migration with β_1 , which measures the difference in the earnings of migrants and non-migrants, adjusted for differences in the age profile.³¹

We estimate equation (1) from a sample combining all Norwegian-born men between the ages of 38 and 50 from the 100 percent 1900 Norwegian Census and the 1 percent sample of the 1900 US Census. The first column of Table 3 shows that Norwegian migrants to the United States earned 60 log points (82 percent) more than men living in Norway in 1900. Columns 2 through 4 reproduce the OLS estimates from equation 1 in our three matched samples. The implied return to migration in our matched samples ranges from 57 to 64 log points (76 to 89 percent). The population estimate represents the midpoint of this range.

The final three columns of Table 3 use the first matched sample to consider alternative data sources and weighting schemes. The fifth column re-weights our matched sample to reflect the urban status, asset holdings and occupational distribution of fathers in the full population, with little qualitative effect on the results. In column 6, we assign US migrants the average earnings for their occupation from the 1915 Iowa Census (appropriately deflated), which is more

²⁹ 89 percent of men have a recorded occupation in the US or Norwegian Census. In our main matched sample, missing occupation data reduces our sample from 19,970 to 17,758.

³⁰ We also try using the "year of immigration" Census variable to restrict our sample to men who were at least 18 years old at the time of immigration to exclude men who arrived in the US as children. We find qualitatively similar results for the regressions reported in Table 3 and all subsequent tables.

³¹ Although we estimate the earnings gap between the US and Norway at a point in time, we note that the true return to migration is the net present value of potential earnings in the destination country relative to the source over the life cycle.

representative of the urban/rural composition of Norwegian migrants, resulting in a lower return to migration of 55 log points (73 percent). The seventh column builds in the 13 log point wage penalty experienced by Scandinavian migrants in the US labor market as described in section 3a. As expected in this case, the return to migration falls to 47 log points (60 percent). Taken together, these adjustments suggest that the baseline estimates may be overstated due to the native-born and urban bias of the earnings data.

B. Comparing migrant and non-migrant brothers within households

The return to migration estimated in equation 1, β_1 , would be the true return if migrants were randomly selected from the Norwegian population. If, however, migrants are (positively or negatively) self-selected, then β_1 will be biased. We next compare the occupation-based earnings of migrants and their non-migrant brothers to eliminate selection across households. Such selection occurs if men from richer or poorer households are more likely migrate to the US.

Specifically, to eliminate selection across households, we consider the following equation in which the individual error term is decomposed into two components:

$$\ln(Earnings_i) = \alpha + \beta_1'(Migrant_i) + \beta_2'(Age_i) + \beta_3'(Age_i^2) + \alpha_j + v_{ij} \quad (2)$$

where α_j is the component of the error that is shared between brothers in the same household j and v_{ij} is the component that is idiosyncratic to individuals.

Running an OLS regression of equation 2 with household fixed effects will absorb the fixed household portion of the error term (α_j). Such within-household estimation will eliminate bias due to aspects of family background that are correlated both with the probability of

migration and with labor market outcomes later in life.³² In this case, the coefficient β_1' measures the return to migration, free from selection across households.

Table 4 compares the naïve OLS and within-household estimates of the return to migration. In order to contribute to the within-household estimation, a household must contain at least two members who match between 1865 and 1900. We begin in the first row of each panel by conducting OLS on the subsample of households with two or more matched members (including migrant-stayer, migrant-migrant and stayer-stayer pairs). In Match 1, the estimated return to migration for men born in rural areas is 59 log points (80 percent), compared with only 34 log points (41 percent) for men born in urban areas.³³

The second row in each panel adds household fixed effects, which eliminates selection across households. In our baseline matched sample, the main effect is identified from the 449 migrants with a matched brother in Norway in 1900. We also present results from Match 2, which expands our sample to 640 matched pairs. Correcting for selection across households increases the return to migration substantially in the urban sample, to 51 log points in Match 1. In contrast, the estimated return to migration in the rural sample is slightly smaller in this specification. Regressions that apply weights to replicate the marginal frequencies of each characteristic in the general population produce similar results.

³² See Griliches (1979), Altonji and Dunn (1996), Aaronson (1998) and Sacerdote (2004) for examples of within-sibling estimates in other contexts. Ashenfelter and Krueger (1994), Behrman, Rosenzweig and Taubman (1996) and Behrman and Rosenzweig (2002) use pairs of identical twins to estimate the returns to schooling.

³³ The return to migration in this sub-sample is somewhat lower than in the matched samples as a whole, perhaps because households with two matched members are more likely to have a high socio-economic status.

C. Inferring the nature of selection

By comparing the naïve OLS estimate of the return to migration (β_1) and the within-household OLS estimate (β_1'), we can infer the direction and magnitude of selection across households into migration. Specifically, if $\beta_1' > \beta_1$, it would suggest that β_1 was biased downward by negative selection of migrant households. That is, men from the types of households that send migrants to the US would have had low earnings even if they stayed in Norway. In contrast, if $\beta_1' < \beta_1$, it would suggest that β_1 was biased upward by positive selection of migrant households.

We find strong evidence of negative selection across households for migrants leaving urban areas. That is, in the urban sample, we find that $\beta_1' > \beta_1$. In particular, the return to migration estimates increase by 20 to 50 percent (from 34 log points to 51 log points, in Match 1). This pattern suggests that the migration flow from Norwegian cities and towns was drawn from households with either lower average ability, fewer connections, or less wealth. In contrast, the estimated return to migration falls slightly when comparing migrants who originated in rural areas to their brothers who remained in Norway, suggesting, if anything, a pattern of weak positive selection.

Even within households, brothers can differ in unmeasured personal attributes (denoted as v_{ij} in equation 2). Appendix B provides complementary evidence on the return to migration and migrant selection using the gender composition of a man's siblings and his place in the household birth order to instrument for migration. Both of these factors influence a man's expectation of inheriting farmland in Norway, and therefore his probability of migrating to the US. The exclusion restrictions are that these two factors do not affect our measure of occupation-

based earnings directly, and the appendix provides supporting evidence that this was likely the case in our context. We focus on the subsample of men born in rural areas whose childhood household held some assets in 1865. The OLS estimate of the return to migration for this selected sample is 64 log points (90 percent). Our IV estimates range from 67 to 70 log points (95 to 101 percent). The larger IV coefficient suggests that the simple earnings comparison may be biased downward, a pattern that is consistent with mild negative selection.

5. Additional evidence of migrant selection

Thus far, we have inferred the direction of migrant selection by comparing naïve estimates of the return to migration with estimates that eliminate across-household selection. In this section, we provide two forms of supplementary evidence that Norwegian migrants, particularly those from urban areas, were drawn from households at the lower end of the skill distribution.

A. Comparing earnings and wealth of the fathers of migrants and non-migrants

Our first approach is to compare the economic outcomes of the fathers of migrants and stayers in the 1865 Norwegian Census and the 1886 Land Registers. Focusing on fathers' outcomes has two advantages. First, fathers' outcomes are all measured in the same economic environment. Second, fathers' outcomes are pre-determined and not affected by the migration decision. We thus collected information on fathers' occupations, asset holdings and property tax values.

Table 5 compares the occupations, asset holdings, and property tax values of the heads of migrant and non-migrant households in 1865. Households are placed in the "migrant" category if

at least one son lived in the US in 1900, and in the “non-migrant” category otherwise. The first panel includes all households with at least one matched son. However, in the full sample, we may erroneously place true migrant households into the non-migrant category because we cannot determine the migration status of unmatched sons. Therefore, in the second panel, we restrict our attention to households in which all sons are matched either to the US or to Norway in 1900. In both cases, we find evidence of negative selection from urban and rural areas. That is, the heads of households that sent migrants to the US were less likely to own assets and more likely to hold occupations with below-median average earnings.

As expected, the patterns are stronger when we focus on households in which the migration status of all sons is known. In the urban portion of this subsample, we find that the heads of migrant households are 6.1 percentage points less likely than the heads of non-migrant households to hold an occupation with above-median earnings (although this difference is not statistically significant) and 8.0 percentage points less likely to own assets. In rural areas, heads of migrant households are also 5.6 percentage points less likely to hold an above-median occupation and 3.1 percentage points less likely to own assets. In addition, rural heads of migrant households are 2.7 percentage points less likely to match to the Land Register of 1886. Failure to match to the register is a strong indication that the household did not have any taxable assets.³⁴ Conditional on matching to these tax records, we find little difference in the property tax bill between migrant and non-migrant households.

³⁴ Failure to match to the tax records can also occur because of imperfect matching conditions such as mortality between 1865 and 1886 or improperly transcribed names.

B. Examining the performance of migrants in the US labor market

Unlike the previous sections, which compared migrants to men who stayed in Norway, here we ask how Norwegian migrants performed in the US labor market relative to US natives both upon first arrival and after some time spent in the country.

Given that migrants may need some time to assimilate, we estimate the earnings gap between migrants and US natives at various points, allowing migrants to have a sufficient amount of time to acquire US-specific skills. We use the 1% IPUMS samples of the US Census from 1900-1930 to compare Norwegian migrants with US natives.³⁵ In order to follow cohorts over time, we estimate:

$$Y_{ijmt} = \left\{ \begin{array}{l} \alpha_j + \alpha_t + \delta_1 \text{Migrant}_i + \delta_2 \text{Age}_{it} + \delta_3 \text{Age}_{it}^2 + \delta_4 m_i + \\ \delta_5 m_i^2 + \delta_6 \text{Migrant}_i * (t - m_i) + \delta_7 \text{Migrant}_i * (t - m_i)^2 + \delta_8 X_{it} + \varepsilon_{ijmt} \end{array} \right\} \quad (5)$$

where Y_{ijmt} is the occupation-based earnings in year t of individual i who migrated to the US in year m and resides in US state j . We restrict our attention to men who migrated to the US at age 20 or older. The variable $t-m$ is the number of years since the individual immigrated to the US.³⁶ X is a dummy for whether the individual is married. We run the regressions separately for urban and rural areas.

The two panels of Figure 4 illustrate the results from equation 5 for a cohort of migrants from an urban or rural area who were born in Norway in 1875 and who migrated to the US in

³⁵ To make the comparison more meaningful, we exclude the US South, which contained few international migrants.

³⁶ Results are qualitatively similar in specifications that include dummy variables for age and dummies grouped by 3-year ranges for years in the US.

1900 at the age of 25. The figure suggests that Norway-to-US migrants earned lower wages than US natives and experienced little convergence over time.³⁷

This question of the relative performance of migrants at the destination is not only interesting in itself, but can also shed light on migrant selection. Assuming that the average skill level of Norwegian born men is the same as (or higher than) that of US-born men, the finding that migrants' earnings do not converge to natives' is consistent with negative selection.³⁸

6. Conclusion

We know surprisingly little about how migrants during the Age of Mass Migration were selected from the European population and about the economic return from their journey. In this paper, we construct a unique data set of Norwegian-to-US migration to estimate the return to migration in the presence of selection into migration. We compare the earnings of men who moved to the US and their brothers who stayed behind in Norway. This approach eliminates the component of migrant selection that takes place across households. We gather further evidence about the nature of migrant selection by comparing the economic outcomes of fathers of migrants and non-migrants, as well as by contrasting the labor market success of Norwegian migrants to the native-born in the US labor market.

We estimate a return to migration from Norway to the US of around 60 percent, which is substantially smaller than the 200-400 percent return for migration from Mexico to the United States today (Hanson, 2006). The contemporary return to migration may be higher than in the

³⁷ We find a similar pattern in panel samples that follow individuals over time between 1900 and 1920 using Census manuscripts from Ancestry.com (not shown). Panel results confirm that the lack of convergence between Norwegian migrants and US natives is not driven by selective return migration. In the panel, however, we could not separate migrants by rural and urban location or exclude the Southern states.

³⁸ Alternatively, migrants may have been positively selected at the outset but either faced high levels of discrimination or lacked key US-specific skills so, ultimately, were unable to converge with US natives. We also note that if the average skill level in Norway was lower than in the US, we cannot draw inferences about migrant selection from this comparison with the native-born.

past because of the sizeable bureaucratic costs of legal immigration and the comparable costs for the undocumented of evading detection, both of which reduce the supply of immigrants to the country. In the late nineteenth century, the border was open to all prospective migrants and, therefore, the return to migration was relatively low. We note, however, that the decision to migrate in the 19th century (as today) may have entailed other non-pecuniary considerations that would have increased the total return to migration (Bertocchi and Strozzi, 2008). For example, at the time, Norway was under Swedish control and limited its franchise to men with wealth, while the US offered the opportunity (for white men) to participate in the democratic system even to new migrants.

We also find that men who faced poor economic prospects in Norway were more likely to migrate to the United States, particularly among men leaving from urban areas. One result of this negative selection is that the naïve return to migration underestimates the true return by 20 to 50 percent. The fact that migrants to the US appear to have been drawn from the lower end of the occupational distribution is consistent with a standard Roy model of migration, as in Borjas (1987), which predicts that men at the lower end of the occupational distribution would have more to gain by moving from relatively unequal European countries to the New World. Furthermore, the fact that European migrants, when unhindered by entry restrictions, were negatively selected from the sending population may explain why the US eventually closed its border to the free flow of labor in 1924 (Hatton and Williamson, 2006) and why some countries explicitly select for more skilled applicants in their immigration policies today.

Appendix A: Data

Our earnings measure is based on assigning men the mean income earned by members of their occupation. Men living in the United States are matched to income data from the 1901 Cost of Living Survey while men living in Norway are matched to mean income by occupation tabulations for the year 1900 published by Statistics Norway and other sources (Haines and Preston, 1991; *Statistik Aarbog*, 1900; Grytten, 2007). We convert Norwegian wages to real, PPP-adjusted US dollars using the 1900 exchange rate and price levels reported in Grytten (2004).

These sources do not report information on earnings for a few large occupations, including farmers and fisherman and white collar workers in Norway. This appendix explains how we estimate the earnings of men in these occupations.

A. Estimating farmers' income

Standard sources do not report information on earnings for owner-occupier farmers in either the United States or Norway. We follow Mitchell, et al. (1922) in estimating the net earnings of owner-operator farmers from farm revenue and expenditures data. For the United States, we use data on farmers in Minnesota, the most common state of residence in our sample, from the 1900 Census of Agriculture. For Norway, we use data for the total value of farm products for the 1900 harvest found in the 1907 Census of Agriculture (*Jordbruksteljinga*).

Appendix Table A1: Estimated earnings for farmers in the United States

	Statistics per farm
INCOME	
Value of farm products not fed to livestock	\$753
Value of house rent and food/fuel produced on farm and consumed by family	\$200 (*)
<i>Gross earnings</i>	<i>\$953</i>
EXPENDITURES	
Labor, fertilizers	\$98
Feed, seed, threshing	\$75 (^)
Taxes	\$27 (#)
Maintenance charges (building, machinery)	\$62 (+)
<i>Total</i>	<i>\$262</i>
NET EARNINGS	\$691

Notes: (*) = Ratio of rent and food/fuel consumed to value of products sold from Goldenweiser (1916).

(^) = Ratio of feed, seed, and threshing charges relative to labor and fertilizers from Goldenweiser (1916).

(#) = Assume tax rate of 0.6% on total value of farm.

(+) = Assume maintenance charge (depreciation) of 0.05 on buildings and 0.15 on machinery. Values of buildings and machinery reported in 1900 Census of Agriculture.

The 1907 Census of Agriculture reports the total value of farm product, rather than average value per farm. According to the 1900 Census, total farm output in Norway is produced by 133,400 owner-operators, 73,200 farm laborers, 24,500 tenant farmers and 35,800 individuals who report being “farmers and fisherman.” To estimate the earnings of owner-occupiers, we need to subtract the value added by tenant farmers and the composite “farmer and fisherman” category; farm labor is already accounted for on the expenditures side of the ledger. The average farm laborer earned \$185 a year (US \$1900). We assume that, with free mobility, tenant farmers would have earned the same amount as farm laborers (in expectation). Therefore, we subtract \$4.5 million (=24,500 · \$185) from the total value of farm product. Furthermore, we assume that men who report being “farmers and fisherman” earn a subsistence living and eat what they produce. Thus, we divide total farm product less \$4.5 million by the number of owner-operators.

Appendix Table A2: Estimated earnings for farmers in Norway

	Statistics per farmer
INCOME	
Value of farm products	\$397 (+)
Value of house rent and food/fuel produced on farm and consumed by family (not reported)	\$106 (*)
Gross earnings	\$503
EXPENDITURES	
	\$109 (*)
NET EARNINGS	
	\$393

Notes: (+) = Unlike the US Census of Agriculture, the value of farm products is derived from transaction data, rather than farmer estimates. Therefore, we assume that the grain used on the farm to feed livestock is already excluded from the total.

(*) = We assume the same ratios as used for the US calculation.

B. Estimating fishermen’s income

The 1906 Statistics Annual (*Statistik Aarbog*) reports the total value of cod, herring, mackerel, salmon, merlan, lobster and oysters sold in 1900. The 1910 volume *Gages Annuels des Domestiques et Salaires des Ouvriers* indicates that, in deep-sea fishing expeditions, fishermen typically received 35-55 of the catch. We divide this total by the 41,680 fisherman in the 1900 Census.

Appendix Table A3: Estimated earnings for fisherman in Norway

	Statistics per fisherman
Value of products sold	\$416
Share provided to fisherman	\$145-\$228. [We use \$200.]
Value of direct consumption of fish	\$48 (*)
TOTAL INCOME	\$248

Notes: (*) Between 1830-1871, the average family spent 8 percent of their expenditures on fish (Grytten, 2004). The average Norwegian family's income was \$300 (in US \$1900), implying an expenditure of \$24 on fish. The families of fisherman likely ate more fish than the average family. We double this value to \$48.

C. White collar workers

With the exception of primary school teachers, we have yet to locate income data for white collar workers in Norway in 1900 (24 percent of the labor force). At the moment, we assign these workers the relevant income level from the United States deflated by the average Norway-US income gap. If the return to skill was higher in the United States than in Norway, this procedure will understate the total return to migration.

D. Comparing the matched samples to subgroups of the population by cause of match failure

In comparing 1865 characteristics of our matched sample to subgroups of the unmatched population, there are two patterns worth noting. First, men who are unique by name and age in 1865 are substantially more likely than others to live in urban areas and in a household with a higher socio-economic status, regardless of whether they eventually can be matched to 1900. This comparison is consistent with our claim that having an uncommon name is associated with being from a higher-status (more cosmopolitan) background. Second, even among the subset of men who are unique in 1865, men who match to 1900 are more likely to live in a high-status household. This fact is likely due to the fact that many of the causes of match failure, including mortality and age misreporting, are negatively associated with socio-economic status.

Appendix Table A4: Comparing matched sample to subgroups of unmatched population

	(1)	(2)		(3)	(4)	(5)		(6)
	Unique in 1900**	Unique in 1865		Not present in 1900	Unique in 1900	Not unique in 1865		Not present in 1900
		Not unique in 1900				Not unique in 1900		
Urban	0.259 (0.438)	0.292 (0.427)		0.252 (0.434)	0.119 (0.324)	0.090 (0.286)		0.124 (0.330)
Urban Areas								
Assets	0.260 (0.438)	0.208 (0.406)		0.198 (0.398)	0.141 (0.348)	0.128 (0.334)		0.125 (0.331)
Occ. > Median	0.591 (0.491)	0.538 (0.499)		0.537 (0.498)	0.471 (0.499)	0.458 (0.498)		0.437 (0.496)
<i>N</i>	5,309	801		12,233	3,457	12,072		2,829
Rural Areas								
Assets	0.664 (0.472)	0.623 (0.478)		0.669 (0.470)	0.677 (0.424)	0.675 (0.468)		0.682 (0.465)
Occ. > Median	0.607 (0.432)	0.549 (0.497)		0.598 (0.490)	0.573 (0.494)	0.546 (0.497)		0.573 (0.494)
<i>N</i>	15,137	1,938		31,695	25,398	121,396		19,817

Notes: ** Matched sample

Appendix B: IV Estimates

Comparing the within-household and naïve OLS estimates of the return to migration reveals selection in the type of households that sent migrants to the US. However, even within households, brothers differ in unmeasured personal attributes (denoted as ν_{ij} in equation 2). In this section, we turn to a complementary instrumental variables estimation that can address both across- and within-household forms of selection. In particular, we aim to find individual characteristics that are correlated with the propensity to migrate but not otherwise associated with labor market potential.

One factor that may have influenced the decision to migrate was the expectation of inheriting farmland in Norway. Historically, property was only passed to sons. Some regions of Norway also relied on a primogeniture system of inheritance wherein the eldest brother stood to inherit the family assets and the corresponding obligation to care for his aging parents. In this social context, younger brothers, who had to “make their own way” in the world, may have been more likely to migrate to the US.³⁹

In accordance with these historical inheritance practices, we consider two instruments for migration status: the gender composition of a man’s siblings and his place in the household birth order. Conditional on the number of siblings, men with fewer brothers received a larger share of the total inheritance. In areas that practiced primogeniture, elder brothers were more likely than younger brothers to inherit land (again, conditional on total number of siblings).

For the IV analysis, we focus on the subsample of men who were most likely to receive an inheritance in land, namely those who were born in rural areas and whose childhood household held some assets in 1865. Furthermore, to minimize measurement error in household composition, we limit our attention to men whose mothers were young enough for a (near)-complete household structure to be observed in 1865.⁴⁰ Using this sample, we estimate the following first stage equation:

$$Migrant_{ij} = \alpha + \gamma_1(Brothers_j) + \Gamma_1(Siblings_j) + \Gamma_2(Rank_{ij}) + \gamma_2(Age_i) + \gamma_3(Age_i^2) + \varepsilon_{ij} \quad (3)$$

where $Migrant_{ij}$ is equal to one for individual i from household j living in the US in 1900. The variables of interest are the number of brothers in the household ($Brothers_j$) and an individual’s rank among these brothers ($Rank_{ij}$). We include both specifications that include both instruments and ones that use them separately. We also include dummy variables for the number of siblings in the household ($Siblings_j$) and control for a quadratic in age. We note that conditional on the

³⁹ In his detailed social history of migration from western Norway, Gjerde (1985) argues that migration was one solution for younger siblings who were constrained by the “system of primogeniture...[under which] they could be nourished and remain on the farm, but they could not marry until they acquired livelihoods that would sustain new families” (p. 86). See Guinnane (1992) and Wegge (1999) for empirical work on the relationship between inheritance systems and immigration in other European contexts.

⁴⁰ We restrict the sample to men whose mothers were 42 years old or less in 1865. This cut-off was selected according to the following logic: (1) Age at first birth was high in Norway during this period; in the 1865 Census, only 13 percent of women had a (surviving) child by the age of 23. (2) Furthermore, children lived with their parents until their late adolescence; 91 percent lived in their childhood until at least the age of 19. Together, these two facts imply that household structure would be incomplete for only 1.2 percent of 42-year old mothers in 1865 (= 0.13 with child by age 23 x 0.09 who left home by age 19). Our results are robust to increasing the age cut-off to 45.

number of siblings in the household, both the number of brothers and an individual's place in the birth order of sons are random.

The second stage of the IV specification is:

$$\ln(Earnings)_{ij} = \alpha + \gamma_1(Migrant_{ij}) + \Gamma_1(Siblings_j) + \gamma_2(Age_i) + \gamma_3(Age_i^2) + \mu_{ij} \quad (4)$$

The independent variable of interest is $Migrant_{ij}$, which is instrumented with the number of brothers in the household $Brothers_j$ and/or with an individual's rank among these brothers $Rank_{ij}$.

The identifying assumption underlying these instruments is that household composition has no influence on labor market outcomes beyond its effect on the probability of migration. However, men who have few brothers or who are eldest among their brothers may not only benefit directly from receiving an inheritance but may also be the beneficiaries of complementary human capital investments in childhood.⁴¹ This possibility is not borne out in our data: among men who remain in Norway, we find no relationship between the composition of a man's childhood household and his labor market outcomes in 1900. We estimated regressions of $\ln(\text{earnings})$ on either number of brothers or birth order for the subsample of men who lived in Norway in 1900, controlling for number of siblings, age and province of birth. We find no relationship between number of brothers and earnings (coeff. = 0.003, s.e. = 0.008) and a statistically insignificant and *positive* relationship between being third or higher in the birth order of sons and earnings (coeff. = 0.021, s.e. = 0.023).⁴²

Our IV results are presented in Table 6. Panel A reports estimates from the first stage equations. Conditional on the number of siblings in the household, the presence of an additional brother increases an individual's probability of migrating to the United States by 1.6 percentage points (relative to the sample migration rate of 11.9 percent). In the second column, we find that men who rank third, fourth or higher in the son order are 5 to 8 percentage points more likely to migrate than their older brothers. However, we do not observe a difference in migration rates between the first and second born sons, suggesting either that inheritance practices may not have followed a strict form of primogeniture in this period or that second sons regularly replaced first sons who died or were otherwise incapacitated before the family property could change hands. Interestingly, in rural households *without* assets, men with more brothers or who are themselves further down the birth order are somewhat *less* likely to migrate (not shown), which we take as suggestive evidence that the effects of household composition estimated here operate through an inheritance channel.

Panels B and C respectively contain OLS and IV estimates of the return to migration for the rural subsample (equation 4 above). The OLS estimate of the return to migration for this selected sample is 64 log points (90 percent). Our IV estimates in panel C range from 67 to 70 log points (95 to 101 percent). The larger IV coefficient suggests that the simple earnings

⁴¹ On the role of sibling gender on human capital investments, see Butcher and Case (1994) and Garg and Morduch (1998). Black, Devereaux and Salvanes (2005) show that birth order affects labor market outcomes in the modern Norwegian economy.

⁴² We acknowledge that, because younger siblings are more likely to migrate to the US, the selection of who remains in Norway may differ by birth order, thereby weakening the power of our test.

comparison may be biased downward by a small amount, a pattern that is again consistent with mild negative selection of migrants from rural areas.⁴³

**Appendix Table B1:
Birth order and number of brothers as instruments for migration to the US**

	(1)	(2)	(3)
A. First stage	<u>Dependent variable = In US in 1900</u>		
Number of brothers	0.016 (0.006)		0.011 (0.006)
2 nd brother		-0.000 (0.012)	---
3 rd brother		0.047 (0.019)	0.037 (0.019)
4 th or higher brother		0.076 (0.035)	0.058 (0.036)
B. OLS	<u>Dependent variable = ln(earnings in 1900)</u>		
In US in 1900		0.642 (0.019)	
C. IV	<u>Dependent variable = ln(earnings in 1900)</u>		
In US in 1900	0.669 (0.436)	0.696 (0.381)	0.668 (0.338)
Over-ID test (p-value)			0.869
N	4031	4031	4031

Notes: Standard errors are reported in parentheses. The sample includes men in Match 1 who lived in a rural household that had some assets in 1865 and whose mother is 42 years old or younger in 1865. The regressions also include a quadratic in age and dummy variables for total number of siblings in the household (see equation 3 in the text). In column 3, we report the p-value from a Sargan (chi-squared) test of over-identification.

⁴³ In the case of a discrete regressor, the measurement error is non-classical. Therefore, the fact that the IV estimate is larger than OLS is not necessarily due to a correction for attenuation bias (see Cameron, 2007).

References

- Aaronson, Daniel, "Using Sibling Data to Estimate the Impact of Neighborhoods on Children's Educational Outcomes," *Journal of Human Resources*, 33 (1998), 915-946.
- Abramitzky, Ran and Fabio Braggion, "Migration and Human Capital: Self Selection of Indentured Servants to the Americas," *Journal of Economic History*, 66 (2006), 882-905.
- Abramitzky, Ran, "The Effect of Redistribution on Migration: Evidence from the Israeli Kibbutz," *Journal of Public Economics*, 93 (2009), 498-511.
- Akee, Randall K. Q., "Who Leaves and Who Returns? Deciphering Immigrant Self-Selection from a Developing Country," *Economic Development and Cultural Change*, 58 (2010) 323-344.
- Altonji, Joseph P. and Thomas A. Dunn, "Using Siblings to Estimate the Effect of School Quality on Wages," *Review of Economics and Statistics*, 78 (1996), 665-671.
- Armstrong, Alexander and Frank D. Lewis, "Capital Constraints and European Migration to Canada: Evidence from the 1920s Passenger Lists." Manuscript, 2009.
- Ashenfelter, Orley and Alan Krueger, "Estimates of the Economic Return to Schooling from a New Sample of Twins," *American Economic Review*, 84 (1994), 1157-1173.
- Atack, Jeremy and Fred Bateman, "'Matchmaker, Matchmaker, Make Me a Match': A General Personal Computer-Based Matching Program for Historical Research," *Historical Methods*, 25 (1992), 53-65.
- Bertocchi, Graziella and Chiara Strozzi, "International Migration and the Role of Institutions," *Public Choice*, 137:1-2, (2008), 81-102.
- Black, Sandra, Paul Devereaux and Kjell Salvanes, "The More the Merrier? The Effect of Family Size and Birth Order on Children's Education," *Quarterly Journal of Economics*, 120:2, (2005), 669-700.
- Butcher, Kristin and Anne Case, "The Effect of Sibling Composition on Women's Education and Earnings," *Quarterly Journal of Economics*, (1994), 531-563.
- "British Pound-Norway Krone Exchange Rate." *Global Financial Data*. <<http://www.globalfinancialdata.com/>>
- Borger, Scott, "Self-Selection and Liquidity Constraints in Different Migration Cost Regimes," Manuscript, 2009.
- Borjas, George, "Self-Selection and the Earnings of Immigrants," *American Economic Review*, 77 (1987), 531-553.

-----, "Immigration and Self-Selection," in *Immigration, Trade and the Labor Market*, John M. Abowd, and Richard B. Freeman, eds. (Chicago: University of Chicago Press, 1991).

-----, "Labor Outflows and Labor Inflows in Puerto Rico," *Journal of Human Capital*, 2:1 (2008), 32-68.

Cameron, A. Colin, "Microeconometrics: Current Methods and Some Recent Developments," in *Palgrave Handbook of Econometrics*, Terence C. Mills and Kerry Paterson, eds. (New York: Palgrave MacMillan, 2009).

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

Chiswick Barry, "Are Immigrants Favorably Self-Selected?" *American Economic Review* 89 (1999), 181-185.

-----, "Are Immigrants Favorably Self-Selected? An Economic Analysis," in *Migration Theory: Talking Across Disciplines*, Caroline Brettell, and James Frank Hollifield, eds. (New York: Routledge, 2000).

Chiswick, Barry and Timothy Hatton, "International Migration and the Integration of the Labor Markets," in *Globalization in Historical Perspective*, Michael D. Bordo, Alan M. Taylor, and Jeffrey G. Williamson, eds. (Boston: National Bureau of Economic Research, 2003).

Feliciano, Cynthia, "Educational Selectivity in US Immigration: How Do Immigrants Compare to Those Left Behind?" *Demography*, 42 (2005), 131-152.

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

Ferrie, Joseph, "A New Sample of Males Linked from the Public Use Micro Sample of the 1850 U.S. Federal Census of Population to the 1860 U.S. Federal Census Manuscript Schedules," *Historical Methods*, 29 (1996), 141-156.

Ferrie, Joseph and Jason Long. "Intergenerational Occupational Mobility in Britain and the US Since 1850." *American Economic Review*, Forthcoming.

Fryer, Roland G. and Steven D. Levitt, "The Causes and Consequences of Distinctively Black Names," *Quarterly Journal of Economics*, 119:3 (2004), 767-805.

Garg, Ashish and Jonathan Morduch, "Sibling Rivalry and the Gender Gap: Evidence from Child Health Outcomes in Ghana," *Journal of Population Economics* (1998).

"Genealogy, Family Trees and Family History Records online." *Ancestry.com*. The Generations Network, Inc. < <http://www.ancestry.com/> >

Gjerde, Jon, *From Peasants to Farmers: The Migration from Balestrand, Norway to the Upper Middle West*, (Cambridge: Cambridge University Press, 1985).

Goldenweiser, E. A., "The Farmer's Income," *American Economic Review*, 6 (1916), 42-48.

Goldin, Claudia and Lawrence F. Katz, "The Returns to Skill across the Twentieth Century United States," NBER Working Paper No. 7126, 1999.

Gould, Eric D. and Omer Moav, "When is 'Too Much' Inequality Not Enough? The Selection of Israeli Emigrants." Manuscript, 2010.

Green, Alan G. and David A. Green, "Immigration and Canada's Wage Structure in the First Half of the Twentieth Century," Manuscript, 2008.

Griliches, Zvi, "Sibling Models and Data in Economics: Beginnings of a Survey." *Journal of Political Economy*, 87 (1979), S37-S64.

Grogger, Jeffrey and Gordon H. Hanson, "Income Maximization and the Selection and Sorting of International Migrants," NBER Working Paper No. 13821, 2008.

Grytten, Ola H., "A Consumer Price Index for Norway, 1516-2003." In *Historical Monetary Statistics for Norway, 1819-2003*. Øyvind Eitheim, Jan T. Klovland and Jan F. Qvigstad, eds. Norges Banks Occasional Papers no. 35, 2004.

Grytten, Ola H., "Nominal Wages in Norway 1726-1940 by Occupation," In *Historical Monetary Statistics for Norway – Part II*. Øyvind Eitheim, Jan T. Klovland and Jan F. Qvigstad, eds. Norges Banks Occasional Papers no. 38, 2007.

Guinnane, Timothy W., "Intergenerational Transfers, Emigration, and the Rural Irish Household System," *Explorations in Economic History*, 29 (1992), 456-476.

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

Hanson, Gordon, "Illegal Migration from Mexico to the United States," *Journal of Economic Literature*, 44 (2006), 869-924.

Hatton, Timothy and Jeffrey Williamson, "What Drove the Mass Migration from Europe in the Late Nineteenth Century?" *Population and Development Review*, 20 (1994), 533-559.

-----, *The Age of Mass Migration: Causes and Economic Impact*. (Oxford: Oxford University Press, 1998).

-----, "International Migration in the Long-Run: Positive Selection, Negative Selection and Policy Williamson and Hatton," In *Labor Mobility and the World Economy*. Rolf J. Langhammer and Federic Foders, eds. (Berlin: Springer, 2006).

- Hvidt, Kristian, *Flight to America*. (New York: Academic Press, 1975).
- Keeling, Drew, "The Transportation Revolution and Transatlantic Migration, 1850-1914," *Research in Economic History*, 19 (1999), 39-74.
- Margo Robert, *Race and Schooling in the South, 1880-1950: An Economic History*. (Chicago: The University of Chicago Press, 1990).
- McKenzie, David and John Gibson, "The Microeconomic Determinants of Emigration and Return Migration of the Best and Brightest: Evidence from the Pacific," *Journal of Development Economics*, forthcoming.
- McKenzie, David, John Gibson and Steven Stillman, "How Important is Selection? Experimental vs. Non-experimental Measures of the Income Gains from Migration," *Journal of the European Economics Association*, forthcoming.
- McKenzie, David and Hillel Rapoport, "Self-selection Patterns in Mexico-US Migration: The Role of Migration Networks," *Review of Economics and Statistics*, forthcoming.
- Mitchell, Wesley, Willford I. King, Frederick R. Macaulay and Oswald W. Knauth, *Income in the United States, Its Amount and Distribution, 1909-1919*. (New York: National Bureau of Economic Research, 1922).
- North Atlantic Population Project and Minnesota Population Center. *NAPP: Complete Count Microdata*. NAPP Version 2.0 [computer files]. Minneapolis, MN: Minnesota Population Center, 2008. [<http://www.nappdata.org>]
- Norwegian Historical Data Centre. *Land Registers 1886*. [computer files]. Troms, Norway: University of Troms, 2010. [http://www.rhd.uit.no/matrikkel/excel_e.html]
- O'Rourke, Kevin and Jeffrey Williamson, "Open Economy Forces and Late 19th Century Swedish Catch-up: A Quantitative Accounting," *Scandinavian Economic History Review* 43 (1995), 171-203.
- O'Rourke, Kevin and Jeffrey Williamson, *Globalization and History: The Evolution of a 19th Century Atlantic Economy* (Cambridge, MA, MIT Press, 1999).
- O'Rourke, Kevin, "The Era of Free Migration: Lessons for Today," IIS Discussion Paper No. 18 (2004).
- Parkerson, Donald H., "Comments on the Underenumeration of the US Census, 1850-1880," *Social Science History*, 15:4 (1991), 509-515.
- Piketty, Thomas and Emmanuel Saez, "Income Inequality in the United States, 1913-1998," *Quarterly Journal of Economics*, 118:1 (2003), 1-39.

Roy, A.D., "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3 (1951), 135-146.

Ruggles, Stephen, et al., *Integrated Public Use Microdata Series: Version 4.0*. Minneapolis, MN: Minnesota Population Center, 2008.

Sacerdote, Bruce, "How Large Are The Effects From Changes In Family Environment? A Study of Korean American Adoptees," *Quarterly Journal of Economics*, 122 (2007), 119-157.

Semmingsen, Ingrid, *Norway to America: A History of the Migration*. (Minneapolis: University of Minnesota Press, 1978).

Soltow, Lee, *Toward Income Equality in Norway* (Madison: University of Wisconsin Press, 1965).

Statistiske Centralbureau, *Arbeidslønninger (Gages Annuels des Domestiques et Salaires des Ouvriers)*. (Kristiania: I Kommission Hos H. Aschehough & Co. 1900, 1910).

Statistiske Centralbureau, *Jordbrukstaellingen I Kongeriket Norge*. (Kristiania: I Kommission Hos H. Aschehough & Co. 1909).

Statistiske Centralbureau, *Statistik Aarbog for Kongeriget Norge (Annuaire Statistique de la Norvege)*. (Kristiania: I Kommission Hos H. Aschehough & Co. 1905.)

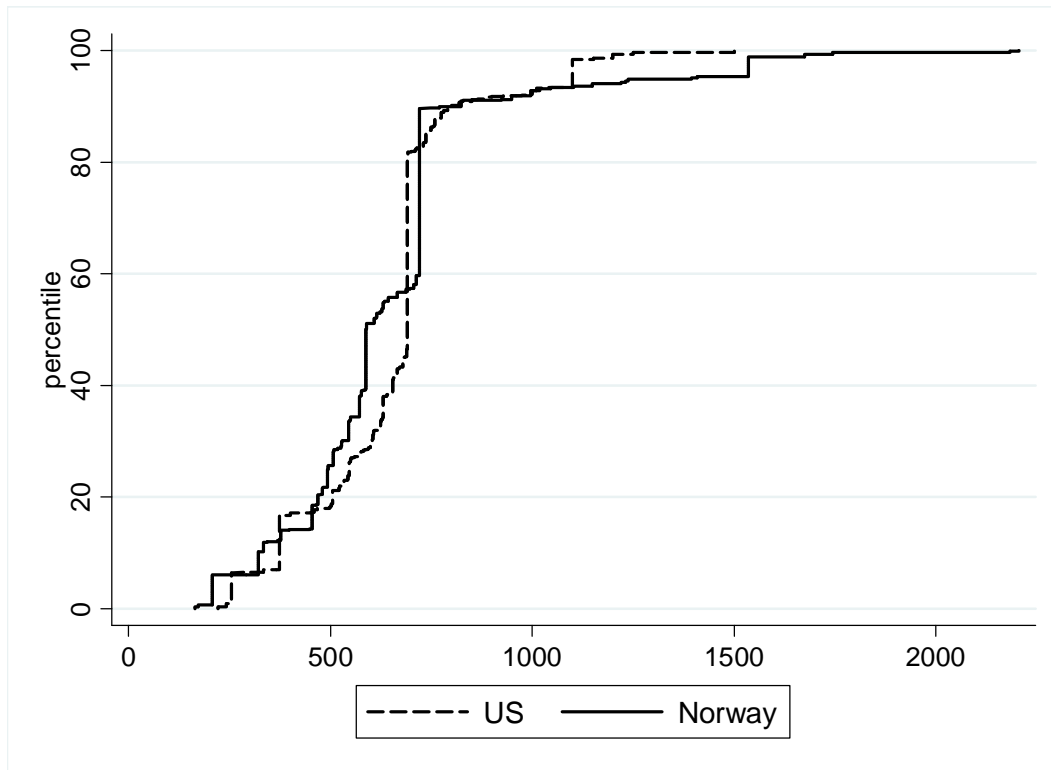
US Bureau of the Census, *Census of Agriculture*. (Washington, DC: Government Printing Office, 1900).

Wegge, Simone A., "To Part or Not to Part: Emigration and Inheritance Institutions in Nineteenth-Century Hesse-Cassel," *Explorations in Economic History*, 36 (1999), 30-55.

-----, "Occupational Self-selection of Nineteenth-Century German Emigrants: Evidence from the Principality of Hesse-Cassel," *European Review of Economic History*, 6 (2002), 365-394.

-----, "Who Migrated and Who Stayed Home? The Village versus the Individual. Self-Selection of Migrants Out of the Home Population in 19th Century Germany." Manuscript, 2010.

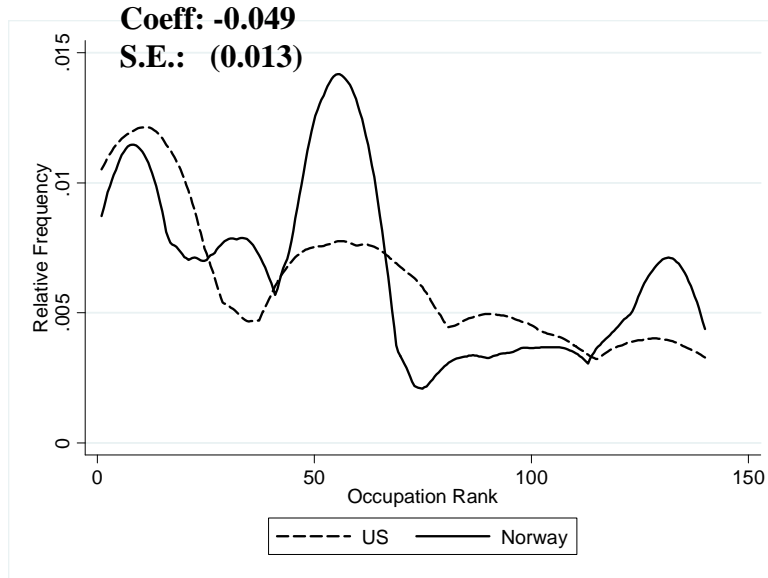
Figure 1: Income distribution in US and Norway in 1900



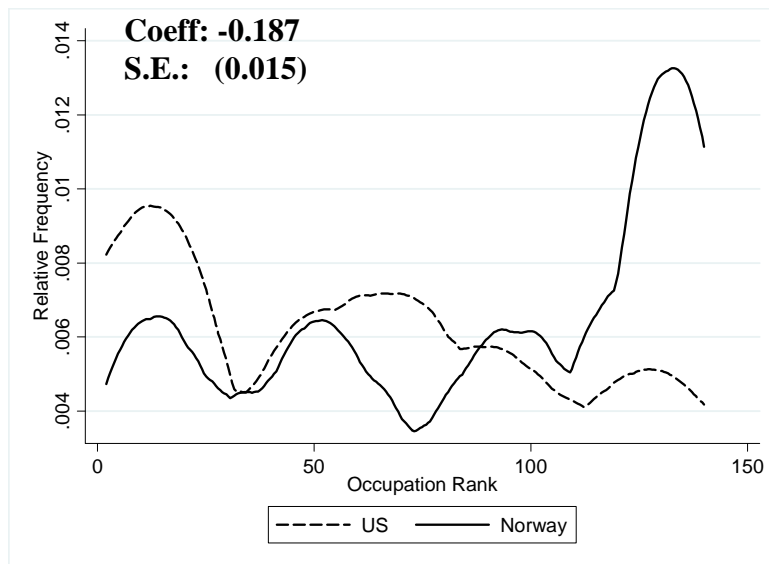
Notes: US and Norwegian distributions contain all men aged 38 to 50 in the respective Censuses of 1900. The X-axis is scaled in 1900 US dollars. Individuals are assigned the mean earnings for their occupation and are arrayed from lowest- to highest-paid occupation. The Norwegian distribution is rescaled to have the same mean as the US distribution (the actual Norwegian and US means are \$350 and \$643 in 1900 US dollars, respectively).

Figure 2: Comparing the occupational distributions of Norwegian-born men in the US and Norway in 1900

A. Born in rural areas

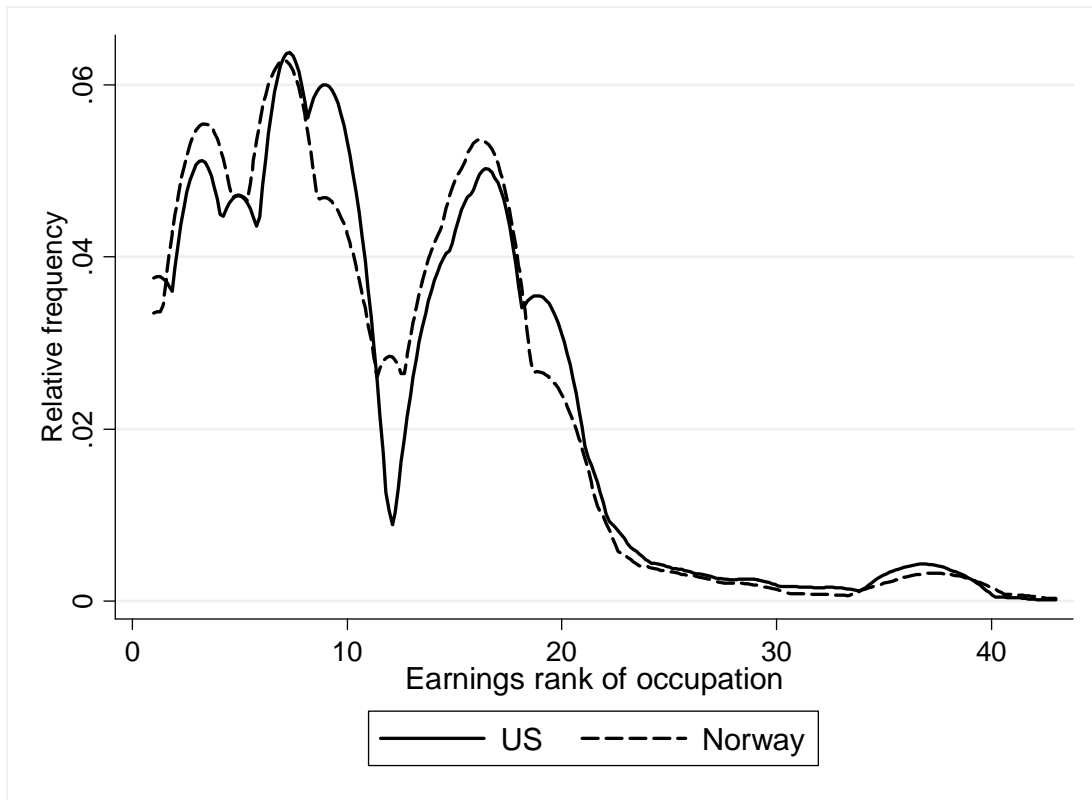


B. Born in urban areas



Notes: Each figure presents the relative frequency of 144 earning categories (representing 189 distinct occupations) for Norwegian-born men in the US and in Norway. All men are assigned the mean US earnings in their occupation. Men are divided by rural or urban place of birth. Farmers, the largest occupational category, is excluded from the figure for reasons of scale. We report coefficients and standard errors from OLS regressions of $\ln(\text{earnings})$ on a dummy for living in the US controlling for a quadratic in age.

Figure 3: Comparing the occupational distributions of migrants who stay in the US or return to Norway between 1880 and 1900



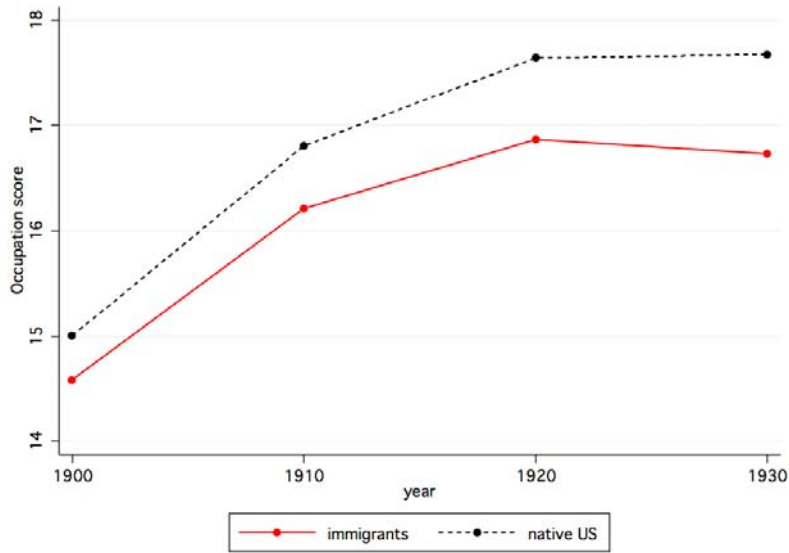
	<u>N</u>	Mean occupation score (1880)	% with occupation score < 12
Unmatched	21,949	14.68	35.20
Matched	3,597	14.65	36.50
- Matched (US)	2,392	14.77	35.50
- Matched (Norway)	1,205	14.39	38.40*

Notes: Return migrants are defined as Norwegian-born men observed in the 1880 US Census who are matched to the 1900 Norwegian census ($N = 1,205$). Persistent migrants are Norwegian-born men in the US Census of 1880 who are matched to the 1900 US Census ($N = 2,392$). For comparison, unmatched men are Norwegian-born men in the 1880 US Census who cannot be matched to either Norway or the US in 1900.

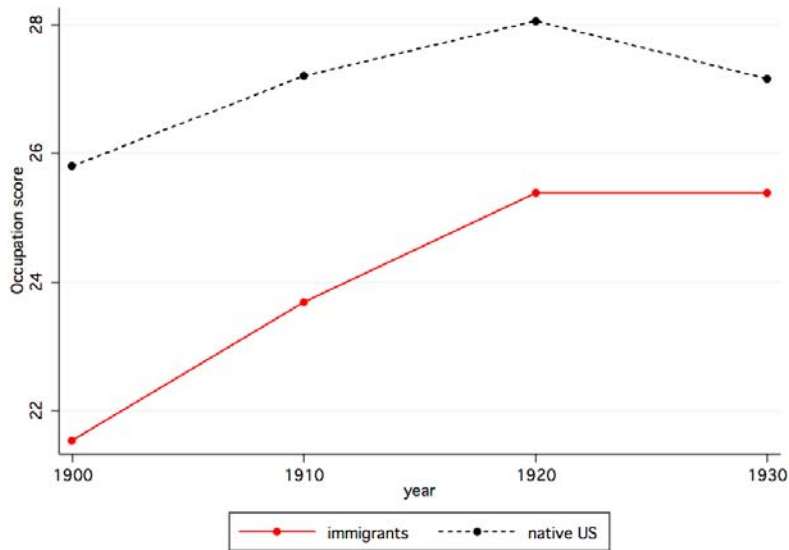
The occupation score measure, which is taken from the 1880 IPUMS sample, is constructed by ordering occupations according to their median earnings in 1950. The mean occupation score and share of the sample with an occupation score in the bottom quartile (score < 12) are reported in the accompanying table. On both measures, the differences between matched and unmatched men are not statistically significant. We mark differences between return and persistent migrants that are statistically different at the 10 percent level with an *.

Figure 4: Assimilation patterns of Norwegian migrants in the US relative to native-born men

A. Rural



B. Urban



Notes: These figures compares the occupation score of Norwegian immigrants who were born in 1875 and moved to the US in 1900 at the age of 20 with the occupation score of US natives (natives from Southern states are excluded).

Table 1:
Common occupations held by Norwegian-born men in the US and Norway

A. Top 10 occupations in matched sample, Norwegian-born men living in the US in 1900

Rank	Occupation	Frequency	Percentage	Earnings
1	Farmers and Planters	1,012	35.81	691
2	Laborers (General)	256	9.05	373
3	Carpenters and Joiners	174	6.15	630
4	Farm Laborers	101	3.57	255
5	Painters, Glaziers, and Varnishers	66	2.33	624
6	Sailors	60	2.12	467
7	Saw and Planing Mill Workers	42	1.49	572
8	Machinists	39	1.38	736
9	Railroad Laborers	36	1.27	460
10	Salesmen	32	1.13	680
Total		1,818	64.33	

N= 2,826. Occupation data collected by hand from Census manuscripts on Ancestry.com. Annual earnings by occupation data from the 1901 Cost of Living Survey reported in Haines and Preston (1991) in year 1900 dollars. Average income of owner-occupier farmers is estimated using data from the US Census of Agriculture.

B. Top 10 occupations in matched sample, Norwegian-born men living in Norway in 1900

Rank	Occupation	Frequency	Percentage	Earnings
1	General Farmers	4,189	22.26	393
2	Farmer and Fisherman	1,522	8.09	321
3	Merchants and Dealers	722	3.84	837
4	Fisherman	709	3.77	248
5	Husbandmen or Cottars	658	3.50	114
6	Farm Workers	597	3.17	175
7	Carpenters	505	2.68	312
8	Shipmasters and Captains	459	2.44	298
9	Cottar and Fisherman	412	2.19	321
10	Seamen	351	1.87	182
Total		10,124	53.79	

N= 18,820. Historical International Standard Classification of Occupations (HISCO) occupation categories. Annual earnings by occupation data from Statistik Aarbog (1900) and Grytten (2007). Values reported in year 1900 dollars. Average incomes of owner-occupier farmers and fishermen are estimated using data from the Norwegian Census of Agriculture.

**Table 2a:
Comparing the matched sample with the Norwegian population in 1865**

	Population	Match	Difference: (Match - Pop)
Urban	0.140 (0.347)	0.259 (0.438)	0.120 (0.003)
A. Urban			
Name frequency	369.06 (1315.99)	7.212 (22.81)	-361.84 (19.27)
<i>Median</i>	<i>11</i>	<i>2</i>	
Age	8.428 (3.691)	8.492 (3.742)	0.063 (0.055)
Number of siblings	5.092 (1.818)	5.215 (1.843)	0.123 (0.027)
Sibling rank	3.466 (1.450)	3.571 (1.489)	0.104 (0.021)
Above median occupation	0.498 (0.500)	0.592 (0.492)	0.104 (0.022)
Assets	0.137 (0.344)	0.260 (0.439)	0.123 (0.005)
B. Rural			
Name frequency	844.53 (2125.90)	15.87 (32.16)	-829.45 (18.30)
<i>Median</i>	<i>121</i>	<i>7</i>	
Age	8.542 (3.657)	8.837 (3.715)	0.295 (0.030)
Number of siblings	5.203 (1.794)	5.232 (1.852)	0.028 (0.014)
Sibling rank	3.599 (1.498)	3.589 (1.492)	-0.009 (0.012)
Above median occupation	0.594 (0.491)	0.607 (0.487)	0.013 (0.004)
Assets	0.630 (0.483)	0.664 (0.473)	0.034 (0.004)

<i>N</i>	236,274	19,970	256,244
----------	---------	--------	---------

Notes: Column 1 contains means and standard deviations (in parentheses) of individual characteristics for the full population between the ages of 3 and 15 in Norway in 1865. Columns 2 reports similar statistics for the matched sample. Column 3 reports coefficients and standard errors for differences between the matched sample and the total Norwegian population. Name frequency counts the number of individuals in the full population (of all ages) with the same first and last name as the respondent. The number of siblings is inclusive of the individual. Oldest siblings have a sibling rank of one.

Table 2b:
Comparing the matched sample with the Norwegian population in 1900 and with Norwegian migrants in the United States in 1900

	Population	Match	Difference: (Match – Pop)
A. In Norway in 1900			
Age	43.811 (3.725)	43.824 (3.724)	0.011 (0.030)
Married	0.864 (0.342)	0.855 (0.351)	-0.009 (0.003)
Children	2.873 (2.486)	2.885 (2.474)	0.012 (0.019)
ln(earnings)	5.773 (0.425)	5.814 (0.451)	0.040 (0.004)
Urban residence	0.270 (0.444)	0.318 (0.465)	0.048 (0.004)
<i>N</i>	139,535	17,432	156,967
B. In US in 1900			
Age	43.386 (3.734)	43.290 (3.755)	-0.095 (0.164)
ln(earnings)	6.384 (0.322)	6.418 (0.313)	0.035 (0.014)
Norwegian name index	1.432 (0.471)	1.478 (0.435)	0.046 (0.014)
<i>N</i>	647	2,538	3,185

Notes: Column 1 contains means and standard deviations (in parentheses) of individual characteristics for the full population of men in Norway between the ages of 38 and 50 in 1900 and for men in this age range born in Norway and living in United States in 1900. The Norwegian data are taken from the full Census (100%), while the US data are drawn from the 1 percent IPUMS sample, thereby accounting for large differences in sample size. Column 2 reports similar statistics for the matched sample. Column 3 reports coefficients and standard errors for differences between the matched sample and the total Norwegian population. The Norwegian name index is equal to the sum of the probabilities that a man is born in Norway conditional on having a given first or last name; the index ranges from zero to two.

Table 3:
OLS regressions of the return to migration from Norway to the US

Dependent variable = ln(earnings)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Population	Match 1	Match 2	Match 3	Match 1		
					Weighted	Iowa data	Add penalty
In US	0.603 (0.015)	0.609 (0.009)	0.644 (0.009)	0.572 (0.015)	0.468 (0.009)	0.641 (0.008)	0.555 (0.010)
<i>N</i>	144,266	17,758	33,641	7,596	14,853	17,601	17,758

Notes: Standard errors are reported in parentheses. All regressions control a quadratic in age. The first column contains a representative sample of the population of Norwegian-born men between the ages of 38-50 in 1900 from the 100 percent 1900 Norwegian Census and 1 percent 1900 US Census sample (IPUMS). Column 2 reports estimates from the first matched sample, which is based on an iterative matching strategy that searches first for an exact match and then for matches in a one- or two-year age band. Column 3 uses the second matched sample, which allows men to match in Norway by name, age, and province of birth. Column 4 reports estimates from the third matched sample, which instead requires that matched observations be unique within a five-year age band. Columns 5 through 7 return to the first matched sample. In column 5, US migrants are assigned earnings from the 1915 Iowa Census (appropriately adjusted for inflation). We lose 157 observations whose occupations do not match to categories in the Iowa Census. In column 6, we reduce the Cost of Living earnings by 13 log points in each occupation based on the earnings penalty for Scandinavian migrants reported in Hatton and Williamson (1994). Column 7 weights the matched sample to reflect the urban status, asset holdings and occupational distribution of fathers in the full population. We lose 2,905 observations because of missing information (primarily missing data on father's occupations).

Table 4:
OLS and within-household estimates of the return to migration
Households with two or more members in the matched sample

Dependent variable = ln(earnings); Coefficient on =1 if migrant

	Rural, 1865	Urban, 1865
A: Match 1		
OLS	0.592 (0.031)	0.339 (0.038)
Within household	0.507 (0.044)	0.511 (0.055)
Weighted within-household	0.529 (0.054)	0.561 (0.080)
<i>N</i>	2,060	1,160
<i>Number of migrant-stayer pairs</i>	184	195
B: Match 2		
OLS	0.644 (0.021)	0.368 (0.035)
Within household	0.573 (0.030)	0.442 (0.047)
Weighted within-household	0.578 (0.032)	0.518 (0.067)
<i>N</i>	5,906	1,486
<i>Number of migrant-stayer pairs</i>	558	359

Notes: Standard errors are reported in parentheses. Each cell contains coefficient estimates from regressions of ln(earnings) on a dummy variable equal to one for individuals living in the United States in 1900. Regressions also include controls for age and age squared. In each panel, the first row conducts an OLS regression for the restricted sample of households that have at least two matched members in the data set. The second row adds household fixed effects and the third row presents coefficients weighted to reflect the urban status, asset holdings and occupational distribution of fathers in the full population.

Table 5:
Economic outcomes of household heads with migrant and non-migrant sons, 1865

	All households			Households with matched sons		
	Migrant	Stayer	Difference	Migrant	Stayer	Difference
A. Urban						
Occ. > median	0.569	0.579	0.011 (0.023)	0.521	0.583	-0.061 (0.049)
Assets	0.223	0.256	-0.033 (0.017)	0.178	0.258	-0.080 (0.029)
<i>N</i>	835	3,667	4,502	197	990	1,291
B. Rural						
Occ. > median	0.601	0.604	-0.003 (0.014)	0.519	0.576	-0.056 (0.030)
Assets	0.640	0.667	-0.026 (0.012)	0.586	0.617	-0.031 (0.026)
Match tax records	0.087	0.119	-0.032 (0.008)	0.089	0.116	-0.027 (0.017)
Property tax bill	\$2.08	\$2.12	-0.032 (0.179)	\$2.08	\$2.19	-0.118 (0.444)
<i>N</i>	1,715/ 149	12,182/ 1,457	13,897/ 1,606	382/ 34	2,824/ 331	3,206/ 362

Notes: Results for Match 1. The first panel includes all households while the second panel considers only households in which all son can be matched to 1900. In each panel, means are reported in columns 1 and 2 and differences between migrant and non-migrant households, along with standard errors, are reported in column 3. Above median occupations are those earning more than \$311 (urban) and \$393 (rural) per year in 1900 US dollars. We assign income levels to household heads using mean Norwegian earnings by occupation in 1900. Assets is an indicator variable equal to one for men who own a business, own land or are master craftsman in an artisanal workshop. Tax records refers to the 1886 Land Registers. For fathers who match to the tax records, we report the value of the property tax bill in 1900 US dollars. The data is provided in *spesiedalers* and *marks*. In 1875, one *spesiedaler* was equal to four Norwegian kroner. Sample sizes refer to the total number of matched households and to the number of matched households that can be found in the tax records, respectively.