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 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, Izi Sin, Gunnar Thorvaldsen, Gui Woolston, Gavin Wright and members of the KALER group at UCLA. We thank seminar participants at Harvard, Queen's, Simon Frasier, Toronto and Yale, as well as conference participants at the Economic History Association, the Social Science History Association and the Development of the American Economy and Labor Studies groups at the NBER. Matthew Baird helped to collect data from Ancestry.com. John Parman and Sula Sarkar generously shared data with us. We acknowledge financial support from the California Center for Population Research and the Center for Economic History at UCLA. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

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
JEL No. J61,N31

ABSTRACT

The Age of Mass Migration (1850-1913) was among the largest migration episodes in history. Unlike today, the United States maintained an open border in this era. We compile a novel dataset of Norway-to-US migrants and estimate the return to migration while accounting for migrant selection. Our first method compares migrants to their brothers who remained in Norway; our second exploits the fact that, under primogeniture, older sons in land-owning families were less likely to migrate. We find that these migrants, unhindered by entry restrictions, were negatively selected from the sending population, and that the 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. First, were migrants positively or negatively selected from the European population? We test whether the US acquired 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. Secondly, what was the economic return to migrating from Europe to the United States in the late nineteenth century?

Understanding migrant selection is of particular importance in this historical period. Given the magnitude of the migration flow, the skill composition of departing migrants had potentially large implications for relative economic growth. If migrants were negatively selected, for example, this outflow may help to explain convergence between the US and poorer European countries. Furthermore, the US maintained a nearly open border in the late nineteenth century, allowing us to study the economic process of migrant self-selection without interference from the bureaucratic factors that govern migrant selection today. In contrast, in the current period, the immigrant flow is a product of both individual migration decisions and the application of

complicated entry rules and restrictions, obscuring the underlying economic forces. Thus, comparing our findings with contemporary studies can illuminate the effect of modern immigration policy on migrant selection.

Our empirical methods are also of general interest to labor economics and the economics of migration. Because migrants may not be randomly selected from the sending population, it is challenging to separately identify the return to migration and the selection into migration. Attempting to measure the return to migration with a naïve comparison of migrants and stayers would be confounded by migrant selection. For example, Europe-to-US migrants may earn more than men who remained in Europe because the brightest people, who would have enjoyed higher earnings even if they had stayed in Europe, are the most likely to move. 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.

We use two alternative and complementary empirical strategies to estimate the return to migration while accounting for selection into migration. First, we use OLS regressions to compare the earnings of migrants to the earnings of their brother(s) who remain in Europe. This within-brother estimate eliminates selection across households, which can result from differing propensities to migrate for households whose members face poor economic opportunities in Europe or for households with the financial capacity to move.

Selection might also occur within households because brothers vary in their innate capacities or social roles. Our second approach addresses the potential for selection both across and within households. We use instrumental variable regressions to compare the earnings of migrants and non-migrants, using birth order to instrument for migration. According to inheritance customs, older brothers were more likely to inherit their parents' land;

correspondingly, we find that younger brothers were more likely to migrate to the US, especially from rural areas and from households that owned land. In some specifications, we also make use of the fact that primogeniture customs were stronger on the western coast of Norway than in the more developed Southeast. In particular, we allow the effect of birth order and its interaction with land ownership to vary by region, thereby controlling for other social or biological aspects of birth order that may be correlated with later labor market outcomes.

Beyond providing more accurate estimates of the return to migration, these methods allow us to infer the nature and extent of migrant selection both across and within households. Specifically, a comparison between the within-brother estimate and the naïve OLS estimate reveals the nature of selection across households, and a comparison between the IV estimate and the within-brother estimate reveals the nature of selection within households.

We focus on Norwegian migrants to the US. 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. More importantly, Norway completely digitized two censuses from the period (1865 and 1900), allowing us to follow large samples of migrants and non-migrants over time in Norway and to the US (Ferrie, 1996). Specifically, 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 to their birth family in Norway in 1865. For comparison, we are also able to follow migrants' brothers and a sample of other men to the Norwegian labor market in 1900.

We observe each individual's occupation in either the US or the Norwegian labor market in 1900.[1] We then assign individuals the mean earnings for their occupation in either Norway or

[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

the US (in real PPP-adjusted 1900 US dollars). For simplicity, we often refer to this occupation-based earnings measure as "earnings," but it can also usefully 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. Despite this drawback, the historical data has an important advantage over its 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 was taken, rendering historical Census data the only large data set available for sibling comparisons or household-based instruments for migration.

We find evidence of negative selection both across- and within-households. That is, men with poorer economic prospects in Norway were more likely to move to the US in the late nineteenth century.[2] Both of our estimation strategies suggest that the return to migration is understated by 20-30 percent due to a process of negative selection. Once accounting for migrants' negative selection, we estimate a return to migration of 60 percent for men born in urban areas and 120 percent for men born in rural areas. Such returns are lower than contemporary estimates for the return to migration from Mexico to the United States (200-400 percent; see Hanson, 2006).

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

---

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] Again, we note that we can only measure selection across occupations, and migrants may have been the brightest and most motivated among the low-skilled.

presents the results of our two estimation strategies – namely conducting a within-brother analysis and using birth order as an instrument for migration. Section 5 concludes.

2. Historical context and related literature

Between 1850 and 1913, more than 40 million Europeans moved to the New World, nearly two-thirds of whom settled in the United States (Hatton and Williamson, 1994).[3] Initially, migrants from the British Isles and Germany constituted the majority of the migrant flow to the US. These early migrants were 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 95 of every thousand Norwegians left the country.

Prospective migrants weighed the cost of passage to the New World against the expectation of higher wages. The shift from sail to steam technology on the Atlantic led to a large reduction in the cost of migration over the nineteenth century. By the 1860s, the average trans-Atlantic voyage cost around $20 for a ticket in steerage and lasted 12 days (Hatton and Williamson, 1998, p. 14; Keeling, 1999). Taking foregone earnings into account, the cost of migration represents around 18 percent of the annual earnings of a Norwegian farm laborer.[4] Friends and family in the destination country helped to defray the cost of passage for new arrivals. In the late nineteenth century, 40 percent of Norwegian migrants travelled on pre-paid steamship tickets financed by friends or relatives (Hvidt, 1975, p. 129).

[3] The first paragraphs of this section are based on Hatton and Williamson (1998) and Chiswick (2003).
[4] 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.

There is little quantitative evidence on the return to migration from Europe to the New World. Hatton and Williamson (1994) proxy for the expected return to migration with the ratio of real wages between source and destination countries. They focus on wages in a few comparable urban occupations in the building trades. In the 1890s and 1900s, a semi-skilled urban laborer in Norway earned around 40 percent of what an equivalent worker in the United States earned. This wage ratio translates into an expected return to migration of 150 percent. However, given that the majority of Norwegian migrants hailed from rural areas, a comparison of wage rates in urban occupations alone is not an ideal measure of the economic return to migration. Moreover, workers in Norwegian cities held a variety of occupations (illustrated in Table 1) and the experience of urban laborers might not be representative. One of this paper's contributions is to estimate the return to migration by comparing a representative sample of actual Norwegian migrants who hold a range of occupations with men who remained in Norway. We estimate much lower returns to migration (between 60 and 120 percent) than suggested by previous literature.

The direction and intensity of migrant selection during the age of mass migration are not well understood. To our knowledge, Wegge (2002) is the only paper to address skill-based migrant selection in the nineteenth century.[5] Wegge finds that members of the richest and poorest occupations in Germany were less likely to migrate than workers in the mid-skill range, such as machinists, metal workers and brewers. Even if the poorest migrants stood to gain the most through migration, they may have lacked the resources necessary to finance their trip.

[5] For a broader discussion of migrant selection from Europe, see Hatton and Williamson (2004). 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.

Migrants who settled in the US in the 1980s and 1990s also appear to have been drawn from the middle of their home countries' skill distributions. Hanson and Chiquiar (2006) document this pattern for the Mexico-US migration by assigning Mexican migrants in the US Census the counterfactual wage that they would have earned if they had remained in Mexico given their education and labor market experience level. Feliciano (2005) finds that migrants have higher education levels than the typical resident of most sending countries, with the notable exception of Puerto Rico. Besides the Puerto Rican case, the patterns observed in contemporary data are driven by a combination of migrant self-selection and selection induced by the bureaucratic process of immigration to the United States.

The Roy model of self-selection predicts that migrants will be positively selected relative to the sending population if either the destination offers a higher return to skill than the source country or migration costs (and borrowing constraints) are sufficiently high (Roy, 1951; Borjas, 1987, 1991).[6] The finding that Mexican migrants are positively selected *despite* the fact that income inequality is higher in Mexico than in the US has lead to an emphasis on the role of migration costs (McKenzie and Rapoport, forthcoming). Hinojosa-Ojeda and McCleery (1992) and Hanson (2006) suggest that the cost of migrating illegally 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.

Contrary to the case of Germany in the mid-nineteenth century or contemporary Mexico, the theory of migrant selection suggests that migrants who left Norway for the US were likely to have been negatively selected. First, in historical terms, the costs of migration were relatively low in this period (18 percent of annual low-skilled earnings in Norway compared to 35 percent of low-skilled earnings in Mexico today). Secondly, while the US wage premium was relatively similar across the skill distribution, low-skilled men could also expect to benefit from substantial

[6] For an alternative view on migrant selectivity, see Chiswick (1999, 2000).

opportunities for occupational mobility in the US . Ferrie and Long (2004) document that only 18 percent of men who held an unskilled, blue collar job in 1850 remained unskilled workers by 1880. Instead, over those three decades, the majority of these men became owner-occupier farmers. We conduct a similar exercise for the 1875 and 1900 Norwegian Censuses and find a quite different pattern. 47 percent of men in unskilled, blue collar occupations in Norway in 1875 remained unskilled workers 25 years later and only 26 percent managed to become farmers.

In addition to our inherent interest in the Age of Mass Migration, we should also point out that our study contributes to the general literature on migrant selection which has been severely limited by data availability. As a result, scholars have either focused on selection on observable characteristics or have turned to unique settings – including the Israeli *kibbutz* (Abramitzky, 2008) or the lottery regulating migration from Tonga to New Zealand (McKenzie, Gibson, and Stillman, forthcoming; McKenzie and Gibson, forthcoming). Historical Census data, which provide information on migrants' names, allow us to create a representative sample for a large and economically significant migration flow.

3. Data and Matching

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

Our goal is to identify Norwegian men in migrant and non-migrant households for whom we can observe labor market outcomes later in life. We rely on three Census sources, the complete digitized Norwegian Censuses of 1865 and 1900 and a data set containing the full population of Norwegian-born men in the US in 1900. We create the full sample of Norwegian immigrants living in the US in 1900 using the genealogy website Ancestry.com. The Norwegian Census data are archived at the North Atlantic Population Project (NAPP).

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 information on wages or income. Instead, we assign men the mean income earned by members of their occupation.[7] 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).[8] The 1901 Cost of Living Survey reports income information for more than 300 occupations in the US. At least one member of our sample is employed in 189 of these categories. We convert Norwegian wages to real, PPP-adjusted US dollars using the 1900 exchange rate and price levels reported in Grytten (2004). The Data Appendix provides more detail on the data sources and assumptions underlying these estimates.

Table 1 reports the ten most common occupations for our sample of matched brothers in Norway and the United States. 42 percent of Norway-to-US migrants in our sample worked in farm occupations, compared to only 26 percent of our sample of Norwegian stayers. Migrants to the US were also far more likely to report being general laborers (8 percent versus 1.4 percent). Other common occupations in both countries include carpenters, fisherman and sawmill operatives.

Our unavoidable reliance on mean earnings by occupation prevents us from measuring the full return to migration. Conceptually, the return to migration can be decomposed into: (1) the presence of higher wages in the US in the typical occupation; (2) the possibility that migrants

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

[8] *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 unemployment is based on reported unemployment for Norwegian migrants in the 1900 US Census.

are able to switch from low-paying to high-paying occupations upon arriving in the US; and (3) the existence of a higher within-occupation return to ability in the US. Our estimate of the return to migration captures only the first two aspects of the total return. We face a related limitation in our ability to describe the extent of migrant selection. Positive selection, for instance, could be generated by either high migration rates among men from occupations with high mean earnings or by high migration rates among men at the 80^{th} or 90^{th} percentile of the wage distribution within their occupation; the reverse is true, of course, for negative selection.

B. Matching Norwegian-born migrants to their birth families

We construct our data set of matched migrant and non-migrant Norwegian households in two steps. First, we match Norwegian-born men from the 1900 US Census to their birth families in the 1865 Norwegian data. After identifying brothers in migrant households, we match these brothers and a comparison group of men from non-migrant households forward to the 1900 Norwegian Census. Our baseline method ("**Match 1**") uses an iterative matching strategy pioneered by Ferrie (1996). We describe this procedure in detail:

(1) We identify 55,079 Norwegian-born men between the ages of 38-50 in the 1900 US Census. Men in this age range were likely to be living with their birth families when they were between the ages of 3-15 in 1865.[9]

(2) We convert all first and last names in this restricted sample and in the complete 1865 Norwegian Census into a phonetic code to address orthographic differences between phonetically equivalent names using the NYSIIS algorithm (see Atack and Bateman,

[9] We restrict our attention to men who are at least three years of age in 1865 to ensure that all observations can match to a two-year age band around the reported age (see step 3). We omit men who are reported to be less than one year of age in the 1865 Census because of concerns about data quality in this subsample.

1992). We focus on the 30,629 of these observations that are unique within their first name, last name and birth year.

(3) We match these unique observations back to 1865 using the following 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 or 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."

(4) For matched observations, we identify all brothers in the household as men with the same relation to household head (usually son).[10] Steps 3 and 4 produce a sample of 3,920 migrants and 6,187 brothers.

(5) Given that our matching process originated with men living in the US in 1900, we have labor market information for all migrants by construction. We search for migrants' brothers and a comparable set of men without matched migrant brothers in the Norwegian Census to find their occupations in 1900. Forward matches are conducted using the iterative procedure described in step 3.[11]

This procedure creates a sample of 4,563 men from migrant households – 3,920 men who lived in the United States in 1900 and 643 who lived in Norway in 1900 – and more than 21,000 men from non-migrant households. Step 3 achieves a backwards match rate of 6 percent among

[10] 97 percent of our pairs are sons of the household head. Grandsons of the household head may be cousins, rather than brothers.

[11] Matches conducted wholly within Norway can make use of an individual's province or municipality of birth in addition to his name and age. Adding an extra matching variable would allow us to distinguish between some men who have the same name and age, thereby increasing our sample size. However, we maintain the more restrictive match by name and age alone for both US-to-Norway and Norwegian matches in order to ensure that migrants and their brothers face the same probability of entering the sample.

all Norwegian-born men living in the US in 1900 or 13 percent among men with a unique name-birth year combination. These rates are comparable to Ferrie's (1996) backwards match rates within the United States of nine and 19 percent respectively. Ferrie's match rates are uniformly higher than ours because he is able to supplement information on name and age with data on state or country of birth.

We are concerned that the iterative nature of this method will produce false matches. False matches may occur because we stop searching once we find an exact match in the iterative process. Thus, we may keep men in the sample who have both an exact match and a "close" match (within a one- or two-year band around the reported birth year).

We design a second matching procedure ("**Match 2**") to address this concern. Match 2 conducts a single match from the United States to Norway (rather than an iterative match) for a restricted sample of men who are unique by name within a *five-year age band* in both locations (two years around the reported age in each direction). In so doing, we limit the potential for false matches in 1900 but we also reduce the ultimate size of the sample. The result is a brothers sample of 1,420 observations – 1,273 men who lived in the United States in 1900 and 147 who lived in Norway – along with nearly 13,000 men from non-migrant households that can be matched between the Norwegian Censuses.

C. Comparing matched samples to the full population

According to aggregate data, 15 percent of the Norwegian population moved to the United States between 1865 and 1900. Our matched samples generate similar out-migration rates of 12-15 percent. Semmingsen (1978, p. 99) reports that over 700,000 Norwegians left the country from 1866 to 1915 on a base population of 2.2 million for a lifetime migration rate of 32

percent. Of these migrants, 75 percent settled in the United States and 64 percent arrived by 1900, implying that 15 percent of Norwegians migrated to the United States by 1900 (= 0.32 emigration rate · 0.75 to US · 0.64 by 1900).[12]

Three aspects of the matching process may lead our sample to be unrepresentative of the general population: first, our matching procedure selects for men with uncommon names; secondly, we will not be able to match men who change their name upon arriving in the US; and finally, we will not capture migrants who returned to Norway or moved elsewhere by 1900. The remainder of this section will consider the existence and magnitude of each source of bias in turn.

Table 2a compares the attributes of men in the two matched samples to the full Norwegian population in the same age range in the 1865 Census. The mean Norwegian man shared his first and last name with 775 others in the country (out of a male population of roughly 600,000). In contrast, members of our matched samples shared their names with fewer than 30 others. Selection on uncommon names produces a sample that is more urban than the typical Norwegian, perhaps because rural families provided their children with a narrower array of given names (Gjerde, 1985, p. 48).[13] While only 14 percent of the Norwegian population lived in an urban area in 1865, 24 percent of Match 1 and 32 percent of Match 2 hail from urban areas.[14] Beyond urban status, men in our matched samples are demographically similar to the population as a whole in terms of age, number of siblings, gender composition of siblings and birth order.

[12] According to Ferenczi and Willcox (1929), 64 percent of the Norwegian migrants to the United States between 1865-1915 arrived by 1900.

[13] The commonness of a man's name is negatively associated with the probability of having been born in an urban area. A regression of birth place on name frequency suggests that 15.4 percent of men with a unique name were born in an urban area, compared to 14.0 percent of men who share their name with 1,000 other Norwegians. This difference is statistically significant.

[14] 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.

Table 2b compares matched migrants to the full population of Norwegian-born men living in the United States in 1900. Perhaps because of their urban origins, migrants in our matched samples settle in counties with a higher urbanization rate (39-40 percent compared to 35 percent for the typical Norwegian migrant) and are somewhat less likely to live in a Norwegian enclave as measured by the share of the population in the county of residence that are Norwegian born.

Our matching procedure will not capture migrants who Anglicize their name upon arrival in the US. Therefore, men in our matched samples are more likely than the typical Norwegian migrant to have a distinctively Norwegian name. Following Fryer and Levitt (2004), we use the complete 1880 US Census to construct indices of a name's distinctively Norwegian content. 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.[15] Men in our matched samples have index values of 1.51-1.60, compared to 1.37 for unmatched Norwegian men in the US in 1900. However, we find no evidence that the "Norwegian-ness" of a man's name is related to our occupation-based earnings measure.[16]

Our sample originates with Norwegian-born men observed in the United States in 1900. According to the aggregate statistics, 25 percent of the Norwegian migration flow eventually returned to Norway (Semmingsen, 1978, p. 20).[17] Return migrants may have been disproportionately drawn from the upper or lower end of the income distribution. On the one

[15] Our first name index is equal to pr(first name | Norwegian-born) / {pr(first name | Norwegian-born) + pr(first name | born elsewhere)} and likewise for our last name index. The full measure adds these two indices together.
[16] We regress ln(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 0.2 between matched and unmatched men would translate into a 0.4 percent difference in earnings which is both small and statistically insignificant.
[17] 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.

hand, migrants who were unsuccessful in the US may have decided to return home. On the other hand, some migrants may practice a "target earnings" strategy, whereby they migrate to a high wage country for a period of time in order to build up saving and return home.

The availability of an intermediate US Census in 1880 allows us to test the selectivity of the return migration flow. 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 1900, one-third of whom 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 1 reveals few discernable differences in the occupational distributions of these two groups.[18] Men who eventually returned to Norway are slightly over-represented at the bottom end of the occupational distribution but the mean occupation score of returners and persisters are statistically indistinguishable from each other.

On net, given these potential sources of bias, men in our matched sample earn slightly more than all men living in Norway or Norwegian-born men living in the US in 1900 (3 to 8 percent). It is important to note that the direction and extent of this bias is comparable in the sending and destination country. Furthermore, these sources of bias will not affect our conclusions about the direction of migrant selection, which are based on a comparison across different estimation methods using the same source of data.

[18] The occupation score variable is contained in the IPUMS data set and is calculated by matching occupations to their median earnings in 1950.

4. Estimating the return to migration and the nature of self-selection

A. The earnings of Norwegian-born men in the US and Norway

Our matched data sets provide a unique opportunity to account for selection bias in the estimated return to migration. However, as a benchmark, we begin by comparing the earnings of all Norwegian-born men living in the United States to the full male population of Norway in 1900.

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, we estimate:

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

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.[19] The US Census data are taken from the Integrated Public-Use Microdata Series (IPUMS).[20]

The coefficient of interest in equation 1 is β_1, which measures the difference in the earnings of migrants and non-migrants. 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. In the next columns, we augment equation 1 with interactions between migration status and age or place of residence. The return to migration increases with both age and years spent in the United States.[21] The average 36 year old migrant earned 53 log points (71 percent) more than his counterpart in Norway, whereas the average 50 year old migrant experienced a return of 65

[19] Over 95 percent of both US and Norwegian observations have a recorded occupation.
[20] 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.
[21] We only report the interactions with age. Results for years spent in the US are qualitatively similar.

log points (91 percent). The pattern of returns by age and years since arrival in the US could reflect assimilation to the US labor market or changing quality of migrant cohorts over time. We also find that the earnings gap between residents of Norway and the United States is larger for those living in rural areas in 1900 (89 percent) than in urban areas (64 percent). This pattern may indicate relative opportunities in rural and urban places or could suggest differential selection from or into rural and urban areas.

Columns 4 through 6 reproduce the OLS estimates from equation 1 for our matched samples. The implied return to migration of 59-62 log points (80-85 percent) is similar to the 82 percent return to migration estimate for the population as a whole. It is also important to note that the earnings data in the population and matched samples are not directly comparable because they rely on different coding schemes in the US – the harmonized occupation codes from the IPUMS or hand-coded occupations from Ancestry.com, respectively.

The 1901 Cost of Living survey, our source of earnings data in the US, may overstate the return to migration for two reasons. First, wages in the industrial cities from which the data were collected may be higher than in the rural communities favored by Norwegian migrants. The fifth column of Table 3 assigns US migrants in the first matched sample the average earnings for their occupation from the 1915 Iowa Census (appropriately deflated), which is more representative of the urban/rural composition of Norwegian migrants. The implied return to migration of 57 log points (77 percent) suggests that the baseline estimates may be overstated by up to five log points due to the data's urban bias. Secondly, native-born workers, who make up a large share of respondents in the Cost of Living Survey, may have earned more than the typical Norwegian migrant. According to worker surveys conducted by the Immigration Commission in the 1900s, Scandinavian migrants earned 15 log points below native-born workers of native parentage in the

same industry due either to discrimination or a lack of US-specific skills (Hatton and Williamson, 1998, p. 146-47). With these two caveats in mind, we caution that our estimates of the return to migration provide an upper bound and may be overstated by as much as 20 log points.

Overall, we find that our matched samples produce similar OLS estimates of the return to migration as in the full population. For the remainder of our empirical analysis, we focus on our matched samples that will allow us to provide better estimates of the return to migration that take migrant selection into account.

B. Comparing migrants and stayers at common skill prices

The higher earnings of Norwegian-born men in the US can be decomposed into three components: the potential to earn higher wages in the typical occupation in the US, the potential for occupational upgrading (or downgrading) in the US, and positive (or negative) selection into migration. As a first step towards separating these explanations, we eliminate the first component by assigning common skill prices to men in our matched samples. In particular, we give all men living in Norway the earnings associated with their occupation in the United States. In essence, we allow all men, both migrants and stayers, to benefit from the higher wages in the US and compare the two groups on the basis of their occupation's rank in the US earnings distribution.[22]

Figure 2 reports coefficients from OLS regressions of earnings on migration status at common skill prices, along with graphical depictions of the full occupational distributions of migrants and stayers. The large return to migration presented in Table 3 disappears for all men at

[22] 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.

these common skill prices and become substantially negative for urban-born men. The negative "return" to migration of 12 log points for the urban sub-sample suggests the presence of either occupational downgrading or negative selection or both.

The figure displays distributions of occupations for Norwegian-born men in the US and Norway arrayed from lowest to highest according to the average US earnings in that occupation. We present the distributions both with and without farmers, the largest occupational category. Men born in rural areas are employed in similar jobs in both countries; if anything, rural-born men living in the US are slightly less likely to be merchants, proprietors or members of other high-paying occupations. By contrast, for men born in urban areas, the occupational distribution in the US is highly skewed toward low-paying jobs like day laborers and servants. The associated negative relationship between earnings and migration is muted by the higher propensity of US migrants to report being owner-occupier farmers, an occupation at the middle-to-upper range of the distribution.

By assigning migrants and stayers common skill prices, we have shown that the presence of higher wages in the US masks the possibility of either downward occupational mobility or negative selection into migration. We pursue more formal tests of the presence of self-selection in the next two sections.

C. Comparing migrant and non-migrant brothers within households

In this section, we explore the role of selection into migration at the level of the household. Migration status may be correlated with unobserved household characteristics. We provide suggestive evidence of this pattern in Table 4, which compares the share of migrants and non-migrants whose fathers were employed in high-, middle-, or low-paying occupations in

1865. Migrants appear to be drawn from lower-skilled households. This pattern is especially strong in urban areas, where only 51 percent of urban migrants had a father in a highly-paid occupation, compared to 57 percent of non-migrants.

More formally, the individual error term (ε_i) in equation 1 can be decomposed into two components: $\alpha_j + \nu_{ij}$, where α_j is shared between siblings in the same household *j* and ν_{ij} is idiosyncratic to individuals. The household component of the error term (α_j) may be correlated with labor market outcomes. Adding household fixed effects to equation 1 will absorb the household portion of the error term. If households that send migrants to the United States are negatively selected on wealth or social connections, then we would expect the within-household estimate to be greater than the OLS coefficient ($\beta^H > \beta_1$) and vice versa if households are positively selected.[23] It is important to note that we cannot interpret β^H as the "true" return to migration if the individual component of the error term (ν_{ij}) is also correlated with earnings. We will address this issue with our IV analysis below.

Table 5 uses our matched samples to compare between- and within-household estimates of the return to migration. In order to contribute to the within-brothers estimation, a household must contain at least two members whose names are unique enough to allow them to match between 1865 and 1900. We begin in the first row of each panel by conducting OLS on this restricted sample. The return to migration in this sub-sample is somewhat lower than in the matched samples as a whole – for example, compare 57 log points (77 percent) in the sub-sample to 62 log points (85 percent) in the full sample for Match 1 – perhaps because households with two matched members are more likely to have a high socio-economic status. Men born in rural areas experienced a substantially higher return to migration than did men born in urban areas

[23] 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.

(columns 2 and 3). In Match 1, the estimated return to migration is 66 log points (93 percent) for men born in rural areas as compared to only 35 log points (42 percent) for men born in urban areas.

The second row in each panel adds household fixed effects. Comparing migrants to their brothers who remain in Norway has little effect on the estimated return to migration in both matched samples. However, the apparent similarity of the coefficients with and without household fixed effects masks a distinctive pattern by place of birth. We find strong evidence of negative selection across households among migrants leaving urban areas. In our urban sample, the estimated return to migration increases by 21 percent in Match 1 and 27 percent in Match 2 when restricting our comparison to brothers in migrant households. In contrast, the estimated return to migration falls slightly when comparing migrants who originated in rural areas to their brothers who remained in Norway. This pattern suggests that the migration flow from Norwegian cities and towns was drawn from households with either lower ability, fewer connections, or less wealth.

From our within-household analysis, we conclude that the direction of migrant selection varied by place of origin. While migrant households originating in rural areas are nearly representative of the population, migrant households in urban areas appear to be negatively selected.

D. Birth order as an instrument for migration

Comparing the migrant status of brothers allows us to assess the type of households that sent migrants to the US. However, even within households, brothers may differ in ability, motivation and willingness to take risks, unmeasured attributes which are captured in the individual component of the error term (v_{ij}). In this section, we turn to a complementary

instrumental variables estimation approach. In particular, we aim to find a personal characteristic that is correlated with the propensity to migrate but is not otherwise associated with labor market potential.

We expect that a man's placement in the birth order will influence his likelihood of leaving Norway for the US. Specifically, older brothers may stand to inherit the family farm or commercial property in Norway and therefore will have less to gain from migration. In his detailed social history of migration from Balestrand county in western Norway, Gjerde (1985) finds 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). Consistent with this qualitative picture, we note that the oldest brother in a household is more likely to remain in his municipality of birth and to be an owner-occupier farmer later in life, two characteristics that may indicate having inherited the family farm.[24] In this social context, we expect younger brothers, who have to "make their own way" in the world, to be more likely to migrate to the US.

Our first stage equation relates the probability of migrating to the United States to a man's rank among his siblings in 1865. We regress:

$$\text{Migrant}_{ij} = \alpha + \gamma_1(\text{sibling rank}_{ij}) + \Gamma_2 A_{ij} + \Gamma_3 F_j + \Gamma_4 B_j + \Gamma_5 P_j + \varepsilon_{ij} \qquad (2)$$

where *sibling rank* is equal to one plus the number of older siblings in the household. We include dummy variables for single year of age (A_{ij}), total number of siblings in the household (F_j), total

[24] In 1900, 27.10 percent of oldest brothers lived in their municipality of birth compared to only 22.76 percent of other men (p-value of difference = 0.00). Similarly, 32 percent of oldest brothers versus 28 percent of other men were owner-occupier farmers (p-value of difference = 0.00). For other empirical work on the relationship between inheritance systems and immigration in other European contexts, see Guinnane (1992) and Wegge (1999).

number of brothers in the household (B_j), and province of residence in 1865 (P_j). γ_1 compares the migration propensities of two men with the same age and the same number of siblings who differ only in their placement in the birth order. [25]

We restrict our attention to men whose mothers were young enough in 1865 for a (near)-complete household structure to be observed in the Census. In particular, we focus on men whose mothers were 42 or younger in 1865; our results are robust to increasing the age cut-off to 45 or 48. We select this cut-off according to the following logic: In the 1865 Census, 13 percent of 23-year old women had a child and 91 percent of children still lived with their parents by the age of 19. Together, these two facts imply that, by age 42, only 1.2 percent of women would have had a child who already left home (=0.13 with child by age 23 · 0.09 who left home by age 19). In order to generate enough power for our instruments, we use our larger matched sample (Match 1) to conduct the IV analysis.

The first panel of Table 6 reports estimates of γ_1 for rural households. Each step down the birth order increases the likelihood of moving to US by 2.8 percentage points (on a base of 14.6). As expected, birth order effects are stronger in families that own land. In landless families, an increase in sibling rank increases the probability of migration by 1.8 percentage points, while in families that own land the migration probability increases by 3.4 percentage points. We find no effect of birth order on the probability of migration for men who lived in urban areas in 1865, which is consistent with primogeniture being practiced primarily in rural areas. Sibling rank and/or rank interacted with land ownership are strong instruments for migration for rural

[25] We also try estimating equation 2 by replacing the "sibling rank" variable with an equivalent measure of birth order among brothers alone. The results are qualitatively similar but we prefer to use sibling rank as our instrument because it increases the statistical power of our first stage regression. We focus on households with six or fewer children but the results are robust to adding larger families.

households. We calculate an F-statistic greater than ten when sibling rank is entered into the equation in a linear fashion (Stock and Yogo, 2005).[26]

One concern with this instrument is that birth order may affect labor market outcomes through social or biological channels other than migration (Black, Deveraux and Salvanes, 2005). We present evidence below showing that there is no relationship between placement in the birth order and later earnings among Norwegian men who either stayed in Norway or migrated to the US. We also address this issue by allowing the effect of birth order and its interaction with land ownership to vary by region. Primogeniture customs were stronger on the western coast of Norway than in the more developed Southeast. It is unlikely that the other mechanisms relating birth order to labor market outcomes, such a parental time with children, followed a similar regional pattern. In the final row of Table 6, we show that the relationship between birth order, land ownership and migration is stronger for the two-thirds of rural households in our sample located to the west of the 20º meridian. In this specification, the comparison between the West and the Southeast allows us to control for other social or biological aspects of birth order that may be correlated with later labor market outcomes.

The second panel of Table 6 presents second stage estimates of the return to migration. The OLS estimate of the return to migration for this selected sample of rural men is 65 log points (92 percent). The corresponding IV estimates range from 75-79 log points (112-120 percent), i.e. an increase of 10-14 log points. The larger IV coefficients suggest that the simple earnings comparison is biased downward by around 20 percent by negative selection. Given that the

[26] In an alternative specification, we replace the continuous variable for sibling rank with a series of birth order dummy variables. We find a positive and monotonic relationship between placement in the birth order and the likelihood of migration. Second and third children are two percentage points more likely to migrate than are first children, though these differences are only marginally significant. Being a fourth, fifth or sixth child increases the likelihood of migration by five, ten or 12 percentage points respectively. The instrument loses power when we include separate dummy variables for each sibling rank.

within-brothers estimates show no evidence of across-household selection in rural areas, most of this selection must occur within households (Table 5). That is, in rural areas, brothers who would have had worse prospects in Norway were more likely to move to the US. One may be concerned that the higher IV coefficients reflect advantages available to older brothers that translate into better labor market outcomes. In the last column of Table 6, we exploit variation in the strength of the institution of primogeniture across regions in Norway. In this case, we find an IV coefficient of 71 log points, which remains notably larger than OLS. We conclude that the pattern of negative selection is robust to choice of instrument.

The observed within-household selection bias of 10-14 log points is large, but not unreasonably so. We find an earnings gap of 43 log points between the highest and lowest earning brothers who both remained in Norway in 1900. If the brother with the lowest earning prospects always migrated to the United States, we would expect a within-household selection bias of 43 log points. Instead, we find a selection bias term of 10-14 log points, suggesting that migrants are disproportionately drawn from lower earning brothers but are not always those with the *lowest* earning prospects. We conclude that, on average men who migrated to the United States would have earned less than their siblings if they had remained in Norway.

The validity of the national instrumental variables procedure (that is, the IV that does not make use of regional variation) rests on the assumption that rank in the birth order affects migration behavior but has no other effect on earnings in the labor market. Table 7 provides evidence consistent with this assumption. We estimate the relationship between log earnings and sibling rank, restricting our attention first to men living in the United States in 1900 and then to men living in Norway. In other words, we are interested in whether being a younger sibling

increases a man's level of human capital sufficiently to increase his labor market outcomes for reasons other than migration behavior.

We find no evidence of an independent effect of birth order on earnings in this period. Each step in the birth order increases earnings by 0.7 percent in Norway while decreasing earnings by -0.9 percent in the US. Both of these estimates are statistically indistinguishable from zero. The fact that the effect of birth order on earnings is small, insignificant and varies in sign across countries suggests that being a younger sibling is not associated with systematically higher earnings. Occupational and geographic patterns by placement in the birth order provide some evidence for why this is the case. On the one hand, older brothers in Norway are more likely to be owner-occupier farmers, a relatively well-paid occupation, while, on the other hand, they are less likely to leave their rural birth place to seek out urban employment opportunities. The lack of a clear birth order effect on earnings is contrary to Black, Deveraux and Salvanes (2005) who estimate a 2 percent reduction in earnings for every increase in sibling rank (from oldest to youngest) in contemporary Norway. The main explanation for this pattern in modern data – access to parental time with children – may not have had a high labor market return in this historical context (see, for example, Price, 2008).

Overall, we find that younger siblings are more likely to migrate to the US. However, for men who stay in Norway, there is no net effect of birth order on earnings in this period. Therefore, we argue that birth order is a valid instrument for migration, one that is particularly strong among families that owned land on the western coast of the country. Larger estimates of the return to migration in an instrumental variables context provide further evidence of migrants' negative selection from the sending population.

5. Conclusion

We know surprisingly little about how migrants during the age of mass migration were selected from the European population and the economic return from their journey. In this paper, we construct a unique data set of Norwegian-to-US migration and use two alternative approaches to estimate the return to migration in the presence of selection into migration. The first method compares 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. Our second approach uses birth order to instrument for migration. Younger brothers in rural areas were more likely to migrate to the US, perhaps because their older siblings inherited the family farm under Norway's system of primogeniture. The IV estimates eliminate selection both across and within households.

We estimate returns to migration from Norway to the US that range from 60 percent for men leaving urban areas to 120 percent for men leaving rural areas. The return to migration in the late nineteenth century appear to be substantially smaller than the 200-400 percent return that Hanson (2006) estimates for the migration from Mexico to the United States today. The contemporary return to migration may be higher than in the 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 were relatively low.

We find that migrant selection generates substantial bias in the naïve estimates of the return to migration. Both across- and within-households, men who faced poor economic prospects in Norway were more likely to migrate to the United States. Due to this negative

selection, the naïve return to migration is underestimated by 20-30 percent. The fact that migrants to the US appear to have been drawn from the lower end of the occupational distribution is consistent with high rates of occupational mobility in the United States relative to Norway in the late nineteenth century (Ferrie and Long, 2004; Ferrie, 2005). A standard Roy model of migration, as in Borjas (1987), predicts that men at the lower end of the occupational distribution would have more to gain by moving to the New World – and that is indeed what we find. The fact that European migrants, when unhindered by entry restrictions, were negatively selected from the sending population, may explain why some countries explicitly select for more skilled applicants in their immigration policies today.

Data Appendix

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*).

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 (*)
Gross earnings	***$953***
EXPENDITURES	
Labor, fertilizers	$98
Feed, seed, threshing	$75 (^)
Taxes	$27 (#)
Maintenance charges (building, machinery)	$62 (+)
Total	***$262***
NET EARNINGS	$691

(*) = 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.

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

(+) = 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.

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

(*) 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.

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.

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.

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), p. 669-700.

"British Pound-Norway Krone Exchange Rate." *Global Financial Data*. <http://www. globalfinancialdata.com/>

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).

----------, "The Economics of Immigration," *Journal of Economic Literature,* 32 (1994), 1667-1717.

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.

----------, "How Ya Gonna Keep 'Em Down on the Farm [When They've Seen Schenectady]?: Rural to Urban Migration in Nineteenth Century America, 1850-70," working paper, Northwestern University, 1999.

----------, "The End of American Exceptionalism? Mobility in the U.S. Since 1850." *Journal of Economic Perspectives*, 19 (2005), 199-215.

Ferrie, Joseph and Jason Long. "Geographic and Occupational Mobility in Britain and the US, 1850-1881." Manuscript, 2004.

"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.

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

Grytten, Ola H., "A Consumer Price Index for Norway, 1516-2003." In *Historical Monetary Statistics for Norway, 1819-2003.* Øyvind Eitrheim, 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 Eitrheim, 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," NBER Working Paper no. 10529, 2004.

Hinojosa-Ojeda, Raul and Robert McCleery, "US-Mexico Interdependence, Social Pacts and Policy Perspectives," in *US-Mexican Relations: Labor Market Interdependence*, eds. Jorge Bustamante, Clark Reynolds and Raul Hinojosa-Ojeda. (Stanford, CA: Stanford University Press, 1992).

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]

Price, Joseph, "Parent-Child Quality Time: Does Birth Order Matter?" *Journal of Human Resources*, 43:1 (2005), p. 240–265.

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).

Statistiske Centralbureau. *Arbeidslonninger* (*Gages Annuels des Domestiques et Salaires des Ourvriers*). (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.)

Stock, James H. and Motohiro Yogo, "Testing for Weak Instruments in Linear IV Regression." in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, Donald W. K. Andrews and James H. Stock, eds. (Cambridge: Cambridge University Press, 2005).

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.

Figure 1: Comparing the occupational distributions of Norway-to-US migrants who either stay in the US or who return to Norway between 1880 and 1900

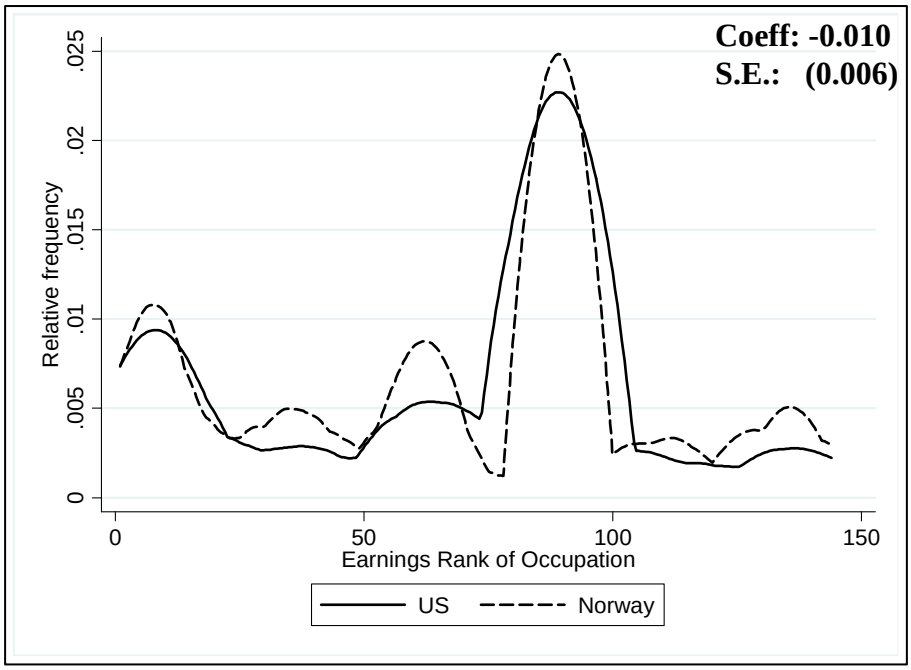
	N	Mean occupation score (1880)	% with occupation score < 12
Unmatched	21,949	14.68	35.20
Matched	3,597	14.65	36.50
- Matched (US)	2,392	14.77	35.50
- Matched (Norway)	1,205	14.39	38.40*

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.

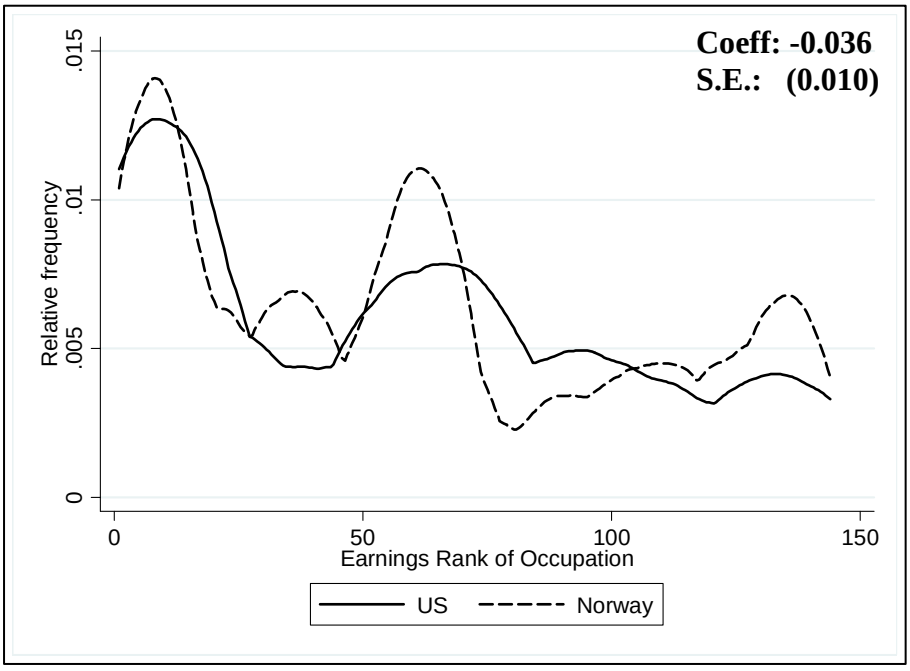
The occupation score measure, which is taken from the 1880 IPUMS sample, is constructed by ordering occupation according to their median earnings in 1950. The mean occupation score and share of the sample with an occupation score in the bottom quartile (score < 12) are reported in the accompanying table. 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 2: Comparing the occupational distributions of Norwegian-born men in the US and Norway in 1900

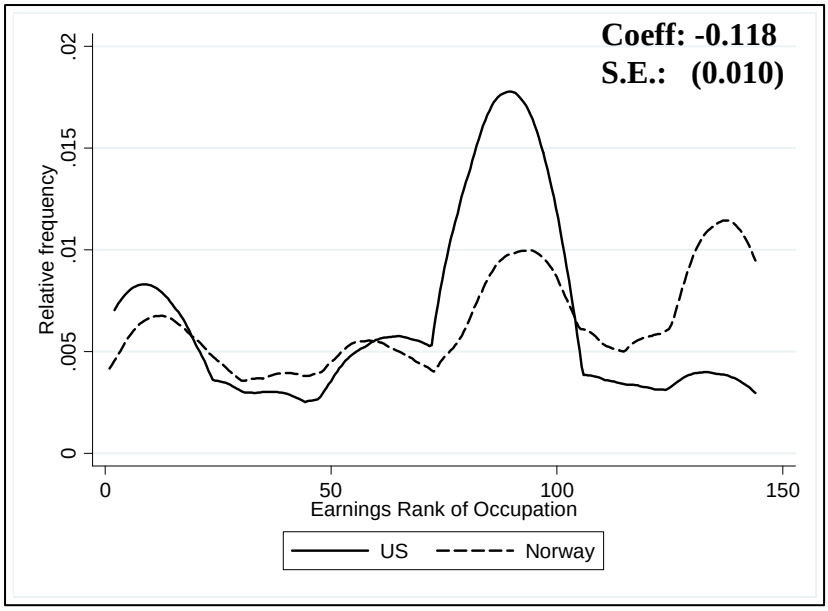
A. <u>Born in rural areas, including farmers</u>



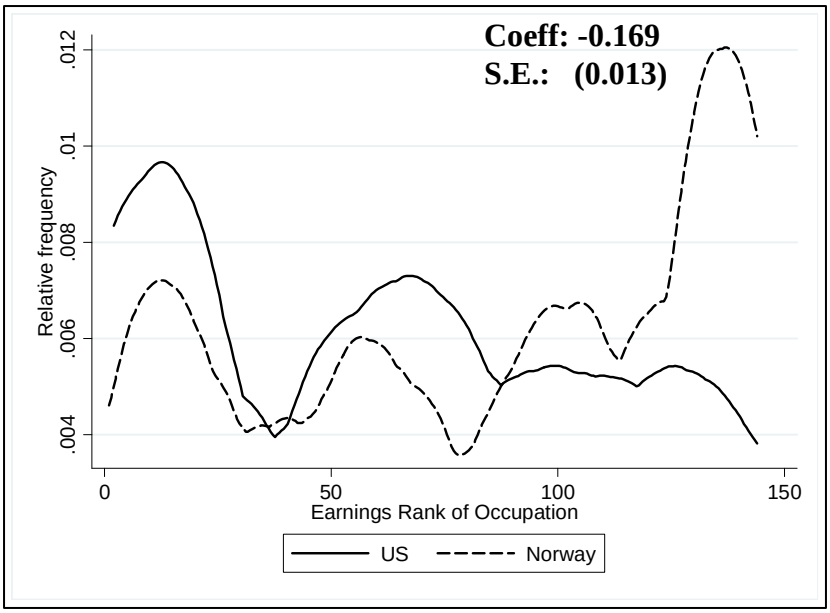
B. <u>Born in rural areas, excluding farmers</u>



C. <u>Born in urban areas, including farmers</u>



D. <u>Born in urban areas, excluding farmers</u>



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 place of birth (rural or urban). Within each place of birth, figures are presented for all occupations (Figures 1a and 1c) or for non-farmers (Figures 1b and 1d). The coefficients and standard errors are from OLS regressions of ln(earnings) on a dummy for living in the US controlling for a quadratic in age.

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	1690	36.99	691
2	Laborers (General)	382	8.32	373
3	Carpenters and Joiners	274	5.97	630
4	Farm Laborers	211	4.6	255
5	Painters, Glaziers, and Varnishers	112	2.44	624
6	Sailors	103	2.24	467
7	Saw and Planing Mill Workers	73	1.59	572
8	Machinists	56	1.22	736
9	Fishermen and Oystermen	55	1.2	550
10	Steam Railroad Laborers	52	1.13	462
Total		3008	65.7	

N= 4591. 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. The Data Appendix provides details on this procedure.

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

Rank	Occupation	Frequency	Percentage	Earnings
1	General Farmers	148	17.11	393
2	Dealer, Merchant, etc.	44	5.09	837
3	Husbandmen or Cottars	28	3.24	113
4	Farmer and Fisherman	27	3.12	321
5	Carpenters	27	3.12	312
6	Farm Workers	27	3.12	175
7	Ship Masters and Captains	16	1.85	297
8	Fishermen	15	1.73	248
9	Sawyers and Sawmill Operatives	14	1.62	269
10	Boot and Shoe Makers and Repairers	14	1.62	276
Total		360	41.62	

N= 865. 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 income of owner-occupier farmers and fisherman is estimated using data from the Norwegian Census of Agriculture. The Data Appendix provides details on this procedure.

Table 2a:
Comparing the two matched samples to the Norwegian population in 1865 and 1900

	Means			Differences	
	Population	Match 1	Match 2	Population & Match 1	Match 1 & Match 2
A. 1865					
Name frequency	775.75	26.89	5.48	-748.85	-21.41
	(2040.00)	(132.40)	(7.07)	(12.94)	(1.27)
Median frequency	*94*	*11*	*3*		
=1 if urban	0.140	0.237	0.316	0.096	0.078
	(0.341)	(0.425)	(0.460)	(0.002)	(0.005)
Age	8.56	8.74	8.77	0.185	0.025
	(3.69)	(3.72)	(3.72)	(0.024)	(0.043)
Number siblings	4.13	4.19	4.25	0.060	0.066
	(1.79)	(1.83)	(1.87)	(0.011)	(0.021)
Sibling rank	2.670	2.708	2.740	0.032	0.036
	(1.530)	(1.570)	(1.610)	(0.010)	(0.018)
Share sisters	0.479	0.479	0.478	0.0001	-0.001
	(0.310)	(0.316)	(0.315)	(0.002)	(0.004)
N	245,765	24,853	10,758	269,618	35,611
B. 1900 in Norway					
ln(earnings)	5.769	5.797	5.857	0.027	0.060
				(0.004)	(0.006)
N	105,057	16,916	7,781	121,973	24,697

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 and 3 reports similar statistics for the two matched samples. Match 1 uses an iterative matching strategy, which searches first for an exact match and then for matches in a one- or two-year age band. Match 2 instead requires that matched observations be unique within a five-year age band. Column 4 reports coefficients and standard errors for differences between the first matched sample and the total Norwegian population and column 5 looks at differences between the two matched samples. The number of siblings is inclusive of the individual. Oldest siblings have a sibling rank of one.

Table 2b:
Comparing the two matched samples to Norwegian migrants in the United States in 1900

	Means			Differences	
	Population	Match 1	Match 2	Population & Match 1	Match 1 & Match 2
Norwegian name index	1.374	1.602	1.510	0.228	-0.129
	(0.485)	(0.371)	(0.424)	(0.008)	(0.015)
ln(earnings)	6.384	6.421	6.433	0.037	0.022
	(0.322)	(0.309)	(0.300)	(0.013)	(0.010)
For county of residence					
Share urban	0.347	0.385	0.402	0.030	0.024
	(0.370)	(0.385)	(0.390)	(0.017)	(0.013)
Share Norwegian	0.071	0.065	0.067	-0.006	0.0017
	(0.063)	(0.064)	(0.066)	(0.003)	(0.002)
N	647	3,461	1,243	4,008	4,704

Notes: Column 1 contains means and standard deviations (in parentheses) of individual characteristics for the full population of Norwegian migrants between the ages of 38 and 50 living in United States in 1900. Columns 2 and 3 report similar statistics for the two matched samples. Match 1 uses an iterative matching strategy, which searches first for an exact match and then for matches in a one- or two-year age band. Match 2 instead requires that matched observations be unique within a five-year age band. Column 4 reports coefficients and standard errors for differences between the first matched sample and the total Norwegian population and column 5 looks at differences between the two matched samples. 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. The share urban is equal to the share of the migrant's county who lives in a town with 2,500 or more residents. The share Norwegian is equal to the share of the migrant's county who were born in Norway.

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

Dependent variable = ln(earnings)

	Full population			Match 1 – COL data	Match 1 – Iowa data	Match 2 – COL data
	(1)	(2)	(3)	(4)	(5)	(6)
=1 if migrant	0.603	0.247	0.635	0.620	0.575	0.586
	(0.015)	(0.151)	(0.017)	(0.008)	(0.008)	(0.014)
Age		0.022				
		(0.006)				
Age · migrant		0.008				
		(0.004)				
=1 if urban			0.143			
			(0.003)			
Urban · migrant			-0.138			
			(0.033)			
N	144,266	144,266	144,266	21,975	21,247	8,769

Notes: Standard errors are reported in parentheses. All regressions control a quadratic in age. The first three columns contain 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). The urban variable in column 3 refers to place of residence in 1900. Columns 4 and 5 reports estimates from the first matched sample 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 6 reports estimates from the second matched sample, which instead requires that matched observations be unique within a five-year age band. In column 4, the US migrants are assigned earnings from the 1901 Cost of Living Survey. In column 5, the US migrants are instead assigned earnings from the 1915 Iowa Census (appropriately adjusted for inflation).

Table 4:
Family background of migrants and non-migrants: Father's income category, 1865

	Migrants	Non-migrants	Difference (p-values)
A. Rural			
High Income	26.23	29.43	-0.031 (0.009)
Middle Income	46.39	44.80	0.016 (0.010)
Low Income	27.37	25.77	0.016 (0.008)
N	*3,065*	*22,627*	
B. Urban			
High Income	51.18	57.43	-0.063 (0.015)
Middle Income	13.68	12.84	0.008 (0.010)
Low Income	35.15	29.73	0.054 (0.014)
N	*1,360*	*6,307*	

Notes: Results for Match 1. We assign income levels to fathers in the 1865 Norwegian Census by matching reported occupations to mean Norwegian earnings in 1900. We convert earnings levels into 1900 US dollars. The high income category refers to fathers above the 80th percentile of this occupation-based earnings measure ($534), middle income is between the 50th and 80th percentiles ($320-$534) and low income is below the 50th percentile ($320).

Table 5:
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

	All	Rural, 1865	Urban, 1865
Panel A: Match 1			
OLS	0.571	0.661	0.351
	(0.017)	(0.021)	(0.029)
Within household	0.552	0.643	0.422
	(0.023)	(0.027)	(0.038)
N	4,571	3,165	1,406
Panel B: Match 2			
OLS	0.491	0.622	0.286
	(0.031)	(0.045)	(0.041)
Within household	0.467	0.578	0.373
	(0.047)	(0.067)	(0.067)
N	1,534	880	654

Notes**:** Standard errors are reported in parentheses. Each cell contains coefficient estimates of log 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. Panel 1 reports results from the first matched sample. Match 1 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. Panel 2 reports results from the second matched sample, which instead requires that matched observations be unique within a five-year age band. 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. The second and third columns conduct similar analyses for men who lived in rural or urban areas respectively in 1865.

Table 6:
IV estimates of the return to migration from Norway to the US
Using sibling rank as instrument for migration

	Coefficients for whole country			Coefficients for western coast
	(1)	(2)	(3)	(4)
A. First stage				
Sibling rank	0.031	0.020	[Dummies]	0.013
	(0.005)	(0.007)		(0.009)
=1 if owns land		-0.067	-0.067	-0.092
		(0.019)	(0.019)	(0.023)
Rank * own		0.017	0.017	0.027
		(0.008)	(0.007)	(0.010)
*F-stat on rank and (rank * own)*	*33.55*	*18.92*	*6.58*	*10.11*
B. Second stage				
=1 if migrant	0.753	0.787	0.754	0.708
	(0.204)	(0.192)	(0.187)	(0.184)

Notes: Standard errors are reported in parentheses. The sample includes men in the first matched sample who live in a rural area in 1865, whose mother is 42 years old or less in 1865, and who have six or fewer siblings (N = 7062). The first panel presents the coefficients of interest from first stage regressions of migration status on sibling rank (1= oldest) and, in some cases, a dummy for household land ownership and an interaction between sibling rank and land ownership. The regressions also include dummy variables for age, number of siblings, number of brothers, and province of residence. The instruments for migration are sibling rank and the interaction between rank and land ownership. In the fourth column, we estimate separate coefficients on sibling rank, land ownership and their interaction for municipalities that are on the western coast or in the Southeast of the country along with a dummy for being on the West coast. We report the interaction terms for the West coast here. Our four instruments in this column are the interactions between region and the two earlier instruments. The second panel contains IV coefficients from regressions of log earnings on a dummy variable for being a migrant and all of the dummy variables in the first stage.

Table 7:
Testing the exclusion restriction:
The relationship between sibling rank and earnings among migrants and stayers

Dependent variable = ln(earnings)

	Norway in 1900		US in 1900	
Sibling rank	0.007	-0.000	-0.009	0.003
	(0.008)	(0.010)	(0.010)	(0.015)
=1 if owns land		-0.019		0.054
		(0.024)		(0.042)
Rank * own		0.011		-0.017
		(0.010)		(0.017)
*F-stat on rank, rank * own*	*0.93*	*1.09*	*0.67*	*0.88*
N	5977	5977	1029	1029

Notes: Standard errors are reported in parentheses. The sample includes men in the first matched sample who live in a rural area in 1865, whose mother is 42 years old or less in 1865, and who have six or fewer siblings. Each cell contains a coefficient estimate from the regression of log earnings on sibling rank (1= oldest) and, in some cases, a dummy for household land ownership and an interaction between sibling rank and land ownership. Regressions also include dummy variables for age, number of siblings, number of brothers and province of residence. The first two columns restrict attention to men who lived in Norway in 1900, while the second two columns consider men who lived in the United States in 1900.