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

Ran Abramitzky
Stanford University

Leah Platt Boustan
UCLA and NBER

Katherine Eriksson
UCLA

[Preliminary and incomplete]

July, 2009

Abstract

Between 1850 and 1913, the United States absorbed nearly 30 million immigrants from Europe. This paper estimates the economic return to migrating from Northern Europe to the United States and the selection of the migrant flow. We construct a novel data set of Norway-to-US migrants and their brothers. Because brothers share a family environment, the earnings of brothers who remained in Norway provide our best estimate for what migrants' earnings *would have been* had they not migrated. A naive comparison of all Norwegian-born men residing in Norway and the US produces returns to migration of 93 percent for those leaving rural areas and 42 percent for those leaving urban areas. Larger within-brother estimates of the returns to migration from urban areas suggest a process of negative selection: households with poorer economic prospects in Norway are more likely to send migrants to the US. An instrumental variables procedure, which uses birth order as an instrument for migration, provides evidence of such negative selection even within households.

* Comments welcome at lboustan@econ.ucla.edu. We have benefited from conversations with Dora Costa, Pascaline Dupas, Joseph Ferrie, Claudia Goldin, Avner Greif, Timothy Guinnane, Lawrence Katz, Naomi Lamoreaux, Robert Margo, Roy Mill, Joel Mokyr, Paul Rhode, Izi Sin, Gui Woolston, Gavin Wright and members of the KALER group at UCLA. We thank seminar participants at Harvard, Yale and the Development of the American Economy group at the NBER for their useful suggestions. Matthew Baird helped us collect data from Ancestry.com and John Parman generously shared data from the 1915 Iowa Census. We acknowledge financial support from the California Center for Population Research and the Center for Economic History at UCLA.

"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

Between 1850 and 1913, an era known as the Age of Mass Migration, the United States received nearly 30 million immigrants from Europe. This paper measures the economic returns to migrating from Northern Europe to the US while accounting for the possibility that migrants may have been either positively or negatively selected from the sending population. In essence, we ask whether the United States acquired the best and the brightest from the European population or whether it absorbed Europe's "tired, poor, huddled masses" who migrated to the United States in search of opportunity.

From the European perspective, migration to the labor-scarce, high wage countries in the New World was an important form of human capital investment. Accurately measuring the associated earnings growth allows us to judge the rate of the return on this investment. From the American perspective, attracting the brightest migrants from Europe may have been an important source of human capital-driven economic growth. Alternatively, if migrants were negatively selected, the US growth record would seem even more impressive for having taken place while the country absorbed the most disadvantaged Europeans.

The fact that migrants may not be randomly selected from the sending population complicates the estimation of the "true" returns to migration. In addition to wage differences

between the two countries, higher earnings among migrants may reflect the fact that migrants are positively selected on ability, motivation or willingness to take risks and therefore would have earned a higher wage regardless of where they lived. In the presence of such positive selection, a naïve comparison of the earning differences between migrants and stayers would overstate the returns to migration. In contrast, migrants may be drawn from households with few resources or social connections in the source country and, due to their family background, may have earned a lower wage regardless of where they lived. If migrants were negatively selected, estimated returns to migration would be understated.

We focus on migration from Norway to the United States, which provides a unique opportunity to measure both the return to migration and the direction and extent of selection of the migrant flow. Specifically, we construct a new data set that allows us to observe Norwegian migrants before and after migrating to the US, and to compare their occupations in the US to both their brothers who remained in Norway and to a broader sample of Norwegian men. We calculate the implied return to migration by assigning men the average earning for their occupation in either the US or Norway.

Our dataset is based on three Census sources, the complete digitized Norwegian Censuses of 1865 and 1900 and the population of Norwegian-born men living in the US in 1900. The data are assembled from the North Atlantic Population Project (NAPP) and the genealogy website Ancestry.com. We match Norwegian-born men from the 1900 US Census to their birth families in the 1865 Norwegian data. We then follow brothers observed in the migrants' households along with a broader sample of Norwegian men to the Norwegian Census of 1900.

Because brothers share a common family environment and a portion of their genetic material, the earnings of brothers who stayed in Norway constitute our best estimate for what

migrants' earnings *would have been* had they too remained. If migrants are positively selected, we expect to find a smaller earnings premium from migration within pairs of brothers than in the whole population. In contrast, if migrants are negatively selected, we expect that the migrant earning premium estimated within brother pairs to be larger than in the whole population.

In a series of naïve OLS regressions conducted with our matched sample, we estimate a 77 percent return to migration, which is a weighted average of a high return for men leaving rural areas (93 percent) and a lower return for men leaving urban areas (42 percent). Our estimate of the return to migration for men leaving rural areas changes little when we compare brothers within the same household, suggesting that migrant-sending households in rural areas were representative of the population. However, in the urban sample, the return to migration within brother pairs is larger than in the population as a whole (52 percent), an indication of negative selection across households.

If brothers share all relevant personal characteristics, our within-household estimates will reveal the "true" return to migration. However, brothers vary in ability. Some families may send the most able or entrepreneurial son to the United States while others may decide to keep their most able son at home to manage the family business or take care of parents in old age. The direction of the bias arising from within-household selection is thus ambiguous. Furthermore, having a brother in the United States may alter the occupational choice of brothers remaining in Norway. Remittances from migrant earnings may be used to finance investments in brothers who stayed in Norway, in which case we will underestimate the return to migration by comparing brother pairs.

Such concerns encourage us to develop a complementary estimation strategy. Broadening our comparison to both migrant and non-migrant households, we use a man's rank in his

household's birth order, as well as an interaction between sibling rank and land ownership, as instruments for migration. Because Norwegian inheritance was based on primogeniture, older brothers were more likely to inherit land and to assume the associated responsibility to care for parents in old age. We find that, conditional on age and family size, younger siblings were more likely to migrate to the United States. Consistent with the role of primogeniture, birth order only affects the probability of migration in rural areas and has a stronger effect on migration among households that own land.

Our first stage regressions reveal that, for men born in rural areas, each step increase in the birth order increases the likelihood of migrating to the United States by 2.8 percentage points. Our second stage regressions produce a higher estimate of the returns to migration than we found when comparing sets of brothers (120 percent compared to 90 percent). Since the instrumental variables procedure controls for both across- and within-household selection, the higher IV estimate suggests the presence of negative selection even within households. That is, the brother who would have had worse economic prospects within Norway is more likely to migrate to the United States. Both of our estimation strategies suggest that the return to migration is understated by 20-30 percent due to a process of negative selection.

The identifying assumption of this IV approach is that being a younger sibling increases the probability of migration but does not directly affect earnings in Norway. We provide supporting evidence for this assumption, demonstrating that there is no significant relationship between a man's rank among his siblings and his labor market outcomes later in life neither among men who remained in Norway nor among Norwegian men who lived in the United States.

The paper proceeds as follows. Section 2 briefly discusses the historical context and related literature on the age of mass migration. Section 3 describes the data and the procedures

we used to match migrants to their birth families in Norway and to match Norwegian men found in the 1865 Census to labor market outcomes later in life. Section 4 presents the results of our two estimation strategies – namely conducting a within-household 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).[1] 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. The average decadal emigration rate from Norway reached 95 per thousand in the 1880s, one of the highest out-migration rates among sending countries in that decade.

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 in the 1860s. In the 1880s and 1890s, the period analyzed in this paper, costs of passage remained roughly constant. Friends and family who were already settled in the destination country helped to defray the cost of passage for new arrivals. In the 1870s, 40 percent of Norwegian migrants travelled on pre-paid steamship tickets financed by friends or relatives.

There is little quantitative evidence on the returns to migrating to the New World in the late nineteenth century. 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

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

comparable urban occupations such as 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 urban wage rates is not an ideal measure of the economic returns 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. By comparing actual Norwegian migrants with men who remained in Norway, we estimate much lower returns to migration (between 50 and 90 percent).

European immigrants in general, and Scandinavian immigrants in particular, tended to be young, single and male. This profile is consistent with a selection process by which prospective migrants with higher net returns to migration were more likely to move to the New World. However, the direction and intensity of migrant selection on the basis of skill during the age of mass migration are not well understood.

To our knowledge, the only paper that has attempted to address migrant selection in the nineteenth century is that of Wegge (2002), which compares migration rates across occupation groups in Germany. She finds that members of the richest and poorest occupations were least likely to migrate, while workers in the mid-skill range, such as machinists, metal workers and brewers, were most likely to do so. Even if the poorest migrants stood to gain the most through migration, they may have lacked the resources necessary to finance their trip. Hanson and Chiquiar (2006) analyze contemporary Mexican migration and find a similar pattern, whereby the middle third of the skill distribution is most likely to move to the US. Other work on migrant selection include Borjas (1987, 1991, 1994), Margo (1990), Chiswick (1999, 2000), Ferrie

(1999), Hatton and Williamson (2004), Feliciano (2005), Abramitzky and Braggion (2006), Abramitzky (2008), and Gould and Moav (2008).

Given the widespread evidence on migrant selection, it is important to explicitly account for selection when estimating returns to migration. This paper does so by constructing a data set in which migrant selection is likely to be less important, namely a data set containing brother pairs and aggregate information on municipality of birth.

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 dataset containing the full population of Norwegian-born men in the US in 1900.[2] The Norwegian Census data are archived at the North Atlantic Population Project (NAPP). We first match Norwegian-born men from the 1900 US Census to their birth families in the 1865 Norwegian data. We then identify all male siblings in a migrant's birth family and match these brothers forward to either the United States or Norway in 1900. For comparison, we also match men in non-migrant households who are in the relevant age range in the 1865 Norwegian Census forward to 1900. The matching procedure is described in the next section.

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

[2] We create the full sample of Norwegian immigrants using the genealogy website Ancestry.com. We compile all cases of men who report Norway as their place of birth from the 1900 Census using a PERL script.

income. Instead, we assign men the mean income earned by members of their occupation.[3] Men living in the United States are matched to income data from the 1901 Cost of Living Survey (Haines and Preston, 1991) while men living in Norway are matched to mean income by occupation tabulations for the year 1900 published by Statistics Norway and other sources (*Statistik Aarbog*, 1900; Grytten, 2007).[4] We convert Norwegian wages to real, PPP-adjusted US dollars using the 1900 exchange rate and price levels reported in Grytten (2004). The 1901 Cost of Living Survey collected information from families in 99 industrial cities. Given that over two-thirds of Norwegian migrants lived in a rural area in 1900, we also try matching our sample to average real earnings by occupation from the 1915 Iowa Census.

Table 1 reports the ten most common occupations for our sample of matched brothers in Norway and the United States.[5] 42 percent of our US sample worked in farm occupations, compared to only 26 percent of our Norwegian sample. 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.

These sources do not report information on earnings for a few large occupations, including farmers and fisherman and white collar workers in Norway. We estimate average income for owner-occupier farmers as farm revenues less expenditures using data from the Censuses of Agriculture in both countries. Many Norwegian fisherman worked on share contracts (*Gages Annual*, 1910). We assign fisherman earnings levels equal to 50 percent of all fish and ocean products caught in Norway in 1900. The Data Appendix provides more detail on

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

[4] *Statistics Norway* reports daily wage rates. We convert these wage rates into annual earnings figures by assuming that Norwegians worked six day work weeks (=313 days) and were unemployed for 0.66 months during the year on average (= 297 days). Our estimate for months spent unemployment is based on reported unemployment for Norwegian migrants in the 1900 US Census.

[5] 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.

the data sources and assumptions underlying these estimates. 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 returns to skill were higher in the United States than in Norway, this procedure will understate the total return to migration.

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 mean wages in the US in the average 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 existence of higher within-occupation returns to ability in the US. Our estimate of the returns to migration captures only the first two aspects of the total returns. 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. Because we are only able to measure between-occupation selection, we will understate the full degree of migrant selection.

B. Matching Norwegian-born migrants to their birth families

We construct our dataset of matched migrant and non-migrant Norwegian households in two steps. First, we identify migrant households by locating Norwegian migrants in their birth families in 1865. Secondly, we match members of both migrant and non-migrant households to

the 1900 Census to observe their labor market outcomes. Our baseline method ("**Match 1**") uses an iterative matching strategy pioneered by Ferrie (1996). We describe this procedure in detail:

(1) We identify 65,961 Norwegian-born men between the ages of 36-50 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 1-15 in 1865.[6]

(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, 1992). We focus on the 37,182 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." To ensure that each observation has the possibility of matching within a two-year age band around the reported birth year, we limit our attention to the 30,629 men who were between the ages of 3-15 in 1865.

[6] We omit men who are reported to be under one year of age in the 1865 Census because of concerns about data quality. For example, men under one year of age reportedly have mothers who are 3.5 years older than men who are exactly one year old (37.5 versus 34 years), leading us to believe that many of these men may have actually been 10 years of age in 1865.

(4) For matched observations, we identify all brothers in the household as men with the same relation to household head (usually son).[7] Steps 3 and 4 produce a sample of 3,920 migrants and 6,187 brothers. Given that our matching process originated with men living in the US in 1900, we have labor market information for all migrants by construction. However, we need to match migrants' brothers forward to 1900 to find their labor market outcomes.

(5) We search for migrants' brothers and all men in non-migrant households in the relevant age range in Norway in 1900. Forward matches are conducted using the iterative procedure described in step 3.[8]

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 – and more than 21,000 men from non-migrant households. Step 3 achieves a backwards match rate of 6 percent among 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.[9] In our forward match, we find 17 percent of the eligible sample.

We are concerned that the iterative nature of this method will produce false matches. Case in point: we construct an implied return migration rate for a sample of migrants who arrived in the US before 1880. We can match 32 percent of these men to the United States in 1900 and 16 percent to Norway. The implied return migration rate in this matched sample (35 percent) is

[7] 97 percent of our pairs are sons of the household head. Grandsons of the household head may be cousins, rather than brothers.

[8] While Norwegian matches could be conducted by name, age and province or municipality of birth, we maintain the more restrictive match by name and age alone in order to ensure that migrants and their brothers face the same probability of entering the sample.

[9] Ferrie's match rates are uniformly higher than ours because he is able to use information on name, age and state of birth for native-born Americans.

higher than the 25 percent return migration rate calculated by the Norwegian Bureau of Statistics for 1880 to 1900 (Semmingsen, 1978, p. 120).[10] 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

Men with uncommon first or last names have a higher probability of being matched between Censuses. As a result, our matched samples are not fully representative of men living in Norway in the late nineteenth century. Table 2a compares the attributes of men in the two matched samples, both to each other and to the full Norwegian population between the ages of 3-15 in 1865. 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, the mean member of Match 1 shares his name with 27 others and the mean member of the more restrictive Match 2 shares his

[10] 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.

name with only five others. Because rural families provided their children with a narrower array of given names, men born in rural areas are less likely to be captured by our matching procedure (Gjerde, 1985, p. 48).[11] 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.[12] Primarily due to their urban status, men in our matched samples who remained in Norway in 1900 earned more than the average Norwegian (by 12 percent in Match 1 and 20 percent for Match 2).

Beyond place of birth, men in the matched samples are demographically similar to each other and to the population as a whole. In 1865, at the average age of 8.7, men in both matched samples were one of 4.2 siblings in the household, 48 percent of whom were sisters.[13] Around a quarter of the sample were the oldest sibling in their household and another quarter were the youngest, with the average sibling rank being 2.7.

Our matched samples may slightly under-represent the probability of leaving Norway, perhaps because some migrants changed their name upon arrival in the United States. Semmingsen, 1978 (p. 99) reports that over 700,000 Norwegians left the country from 1866 to 1915 on a base population (in 1900) of 2.2 million, which implies a lifetime migration rate of 32 percent. 75 percent of migrants leaving Norway settled in the United States and the majority left the country by 1900, suggesting that around 18 percent of Norwegians migrated to the United States by 1900 (= 0.32 emigration rate · 0.75 to US · 0.75 by 1900). We find that 15 percent of Match 1 and 12 percent of Match 2 migrated to the United States over this period.

[11] The commonness of a man's name is negatively associated with the probability of having been born in an urban area. Based on a regression of birth place on name frequency, we estimate 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.

[12] 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.

[13] Older sisters were more likely than older brothers to move out of their birth household at marriage, leading to an underrepresentation of sisters among observed siblings.

Table 2b compares matched migrants to the full population of Norwegian migrants living in the United States in 1900. On the one hand, we are more likely to match men whose names are unique in Norway, a characteristic that is associated with being of higher socio-economic status. On the other hand, we suspect that more successful migrants were more likely to Americanize their first name, thereby making them harder to match. On balance, we find that migrants in our matched samples earn more than the average Norwegian migrant in the same age range (by 4 percent in Match 1 and 6 percent in Match 2). Matched migrants settle in counties with a higher urbanization rate (39-40 percent compared to 35 percent for the typical Norwegian migrant) and are slightly less likely to live in a Norwegian enclave. The typical Norwegian migrant lived in a county whose population was 7.1 percent Norwegian-born – over 10 times the rate in the average county – whereas matched migrants lived in counties in which 6.5 percent of the population was Norwegian-born.

4. Estimating the return to migration

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

Our matched datasets 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. This approach, while not accounting for selection, is commonly used to estimate the return to migration when Census data are available for both sending and receiving country (Hanson, 2006).

Combining all Norwegian-born men between the ages of 38 and 50 from the 100 percent 1900 Norwegian Census in 1900 and the 1 percent sample of the 1900 US Census, we estimate:

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

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

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 returns to migration increase with both age and years spent in the United States. 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 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, changing quality of migrant cohorts over time, or a differential return migration rate among the unsuccessful. 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 returns 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. On the one hand, we may

[14] Over 95 percent of both US and Norwegian observations have a recorded occupation.

expect the return to migration to be lower in the matched samples whose members must be unique by name and age. Having an uncommon name is associated with a higher socio-economic status in Norway and therefore a lower return to moving elsewhere (Table 2). However, 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.

Our source of earnings data in the US, the 1901 Cost of Living survey, may overstate the return to migration for two reasons: first, wages in the industrial cities from which the data was collected may be higher than in the rural communities favored by Norwegian migrants and, secondly, Norwegian migrants may have earned less than others in the sample, particularly the native-born, due to discrimination or a lack of US-specific skills. 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 rather than the 1901 Cost of Living Survey (appropriately deflated). The implied return to migration of 57 log points (77 percent) suggests that the baseline estimates may be overstated by up five log points due to the data's urban bias. Furthermore, Hatton and Williamson (1998, p. 146-47) report that, 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. With these two caveats in mind, we caution that our estimated returns to migration provide an upper bound and may be overstated by as much as 20 log points. However, given that we do not have access to micro data allowing us to calculate migrant-specific earnings by occupation, we will proceed with the data at hand.

B. Comparing migrants and stayers at common skill prices

Estimated returns to migration derived from a simple comparison of Norwegian-born men in the US and Norway can be decomposed into three components: the potential to earn higher wages in every 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 assign common skill prices to all men in the matched samples based on their reported occupation. 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 on their occupation's rank in the US earnings distribution.[15]

Figure 1 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 returns to migration presented in Table 3 disappear for all men at 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.

Figure 1 displays the distribution of occupations arrayed from lowest to highest by average US earnings for Norwegian-born men in the US and Norway. 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 but, in the US, are slightly less likely to be merchants, proprietors and other high-paying occupations. By contrast, for men born in urban

[15] 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.

areas, the occupational distribution in the US is highly skewed toward low-paying jobs like day laborers and servants. The associated negative "return" 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.

C. Comparing migrant and non-migrant brothers within households

By assigning migrants and stayers common skill prices, we have seen that the presence of higher wages in the US masks the possibility of either downward occupational mobility or negative selection into migration. In this section, we explore the role of negative selection by comparing sets of brothers, at least one of whom migrated to the United States. Migration status may be correlated with unobserved household characteristics – such as father's occupation, education, or social connections – that are captured in the error term (ε_i). If, for instance, social connections in Norway was positively correlated with an individual's earnings but negatively correlated with his propensity to migrate, our estimate of β_1 will be biased downward.

The individual error term in equation 1 can be decomposed into two components: $\alpha_j + \nu_{ij}$ where α_j is the component of the error term that is shared between siblings in the same household *j* and ν_{ij} is a component that is idiosyncratic to individuals.[16] To begin with, we assume that the individual component of the error is uncorrelated with migration status. The OLS estimate of the return to migration β_1 will be equal to the true return to migration (β^{true}) plus a bias term due to family background that we will call θ.[17] Adding household fixed effects to equation 1 absorbs the

[16] We find strong empirical evidence that brothers' earnings are more similar than the earnings of two randomly selected men in the Norwegian population. The correlation coefficient between the earnings of two matched brothers who both remain in Norway in 1900 is 0.180, whereas the earnings of two randomly selected men from our matched sample is -0.005.

[17] If we think about the component of the error term that is shared between siblings as a single omitted variable (for example, father's education), θ is equal to the coefficient on father's education from a regression of log earnings on

common portion of the error (α_j), thereby eliminating θ. If households that send migrants to the United States are negative selected on the basis of their family's wealth or social connections, then we would expect the estimate derived from comparing brothers in the same household to be greater than the OLS coefficient ($\beta^H > \beta_1$).[18]

Interpreting the within-household estimate as the "true" return to migration relies on the strong assumption that the individual component of the error term (ν_{ij}) is uncorrelated with migration status. Some models of intra-household allocation of resources predict that households facing a borrowing constraint will invest in the most able child (Garg and Morduch, 1998). Families with limited resources may have sent the most able brother – that is, the brother with the highest ν_{ij} – to the United States. In this case, β^H will be higher than the true return to migration ($\beta^H > \beta^{true}$). The converse will be true if the least able or least favored brother migrates to the US.

Table 4 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 who can be matched 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. The requirement of having two brothers who are unique within name and age likely selects for higher-status families whose members would have had access to better economic opportunities in Norway and therefore a lower return to moving elsewhere.

---

migration status and father's education multiplied by the coefficient from a regression of migration status on father's education.

[18] 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.

We can usefully employ this sample to compare estimates of the return to migration that rely on both within- and between-household variation to those that rely only on variation within households. The second row in each panel adds household fixed effects. In this case, the return to migration is identified from the 20 percent of migrants in our sample who have a brother that we can match forward to 1900 (1071 in Match 1 and 290 in Match 2). We find that comparing migrants to their brothers who remain in Norway has little effect on the estimated return to migration in both samples, suggesting that there is little positive or negative selection in the migrant flow.

However, the apparent similarity of the coefficients with and without household fixed effects in the full sample masks an distinctive pattern by place of birth. The second and third columns of Table 4 subdivides the sample into households that were located in rural and urban areas in 1865. Men born in rural areas experienced substantially higher returns to migration than did their counterparts born in urban areas. 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.[19]

A portion of the differential return to migration for men born in rural and urban areas can be explained by location-specific patterns of selection. We find strong evidence of negative selection out of urban areas while, if anything, our method suggests that migrants from rural areas were positively selected. For example, in Match 1, the estimated return to migration falls slightly (from 93 to 90 percent) when comparing migrants who originated in rural areas to their brothers who remained in Norway. However, among urban men, the estimated return to migration increases from 42 to 52 percent when restricting our comparison to brothers in migrant

[19] The lower return among urban men may partially reflect the lack of accurate data for the wages and salaries of white collar workers in Norway. We find a similar differential in rural and urban returns to migration when excluding farmers and fisherman, two large occupations without accurate wage data.

households. By eliminating the shared household component of the error term, the coefficients increase by 21 percent in Match 1 and by 27 percent in Match 2. This pattern suggests that the migration flow from Norwegian cities and towns appears to have been drawn from households with 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. Migrants originating in rural areas are nearly representative of the population, while migrants from urban areas appear to be negatively selected. We would expect differential selection by region if, for example, the US labor market offered high relative returns to skill in the agricultural sector but low relative returns to skill in urban occupations.[20] Goldin and Katz (2000) present evidence consistent with this pattern from the Iowa Census of 1915, demonstrating that the return to a year of high school education was 11.4 percent for men in farm occupations but only 7.4 percent for men in blue collar occupations.

D. Birth order as an instrument for migration

The difficulty of interpreting within-household estimates when ability varies across brothers prompts us to search for an alternative estimation strategy. In particular, we aim to find an instrumental variable for migration to the US – that is, some 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. *A priori*, the direction of this relationship is not clear. On the one hand, older brothers may stand to inherit the family farm or commercial property in Norway and

[20] This argument assumes that there are sufficiently high training costs preventing migrants from easily switching between the urban and rural sectors.

therefore will have less to gain from migration. In his detailed social history of migration from Balestrand county in western Norway, Gjerde (1985) writes 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). In this case, we expect younger brothers, who have to "make their own way" in the world, to be more likely to migrate to the US. On the other hand, older brothers could convert their inheritance into capital to finance their journey to the New World. Gjerde writes that "children of landholders… were a significant segment of the early emigration from Balestrand [because they were able to use] the capital they received from their inheritance" to finance their move (p. 121). This financial advantage may have grown less important over time as familial and social "networks developed strategies that permitted family members without capital to emigrate" (p. 130).

Our first stage equation relates the probability of migrating to the United States to a man's rank among siblings observed in his childhood household 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 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.

For the instrumental variables analysis, 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. Older mothers are more likely to have had children who already left home by 1865. Using the full Census, we calculate that, in mid-nineteenth century Norway, 91 percent of children still lived with their parents by the age of 19 and only 13 percent of women had a child by the age of 23. Taken together, we estimate that only 1.2 (= 0.09 · 0.13) percent of women who were 42 (=23 + 19) years old or less in 1865 likely had a child who already left home. Therefore, we limit our attention to men whose mothers were 42 or younger in 1865, but our results are robust to increasing the age cut-off to 45 or 48. We also focus on households with six or fewer children but again the results are robust to adding larger families.

The first panel of Table 5 reports estimates of γ_1 for rural households. 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 the fact that primogeniture practices were stronger in rural areas. In rural areas, 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.

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.

Sibling rank and/or rank interacted with land ownership are strong instruments for migration. We calculate an F-statistic greater than ten when sibling rank is entered into the equation in a linear fashion (Stock and Yogo, 2005). The instrument loses power when we include separate dummy variables for each sibling rank.

The second panel of Table 5 presents second stage estimates of the return to migration. Because we use the happenstance of a man's birth order to predict his migration behavior, the IV coefficients should be uncorrelated with either family background (α_j) or individual attributes (ν_{ij}), thereby correcting for both across- and within-household selection. 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 78-83 log points (120-130 percent), i.e. 13-18 log points larger. The larger IV coefficients suggest that the simple earnings comparison is biased downward by around 30 percent by negative selection. Given that the within-brothers estimates show no evidence of across-household selection in rural areas, most of this selection must occur within households (Table 4). That is, in rural areas, brothers who would have had worse prospects in Norway were more likely to move to the US.

The observed within-household selection bias of 13-18 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 13-18 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. We note that our data only allow comparisons across occupations. Thus, while we demonstrate that migrants are drawn from the lower occupational rungs, both across- and within-households, we cannot determine whether it was the most motivated and entrepreneurial among these lower-tier occupations who journeyed to the United States.

The validity of this instrumental variables procedure 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 6 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 returns in the labor market for reasons other than migration behavior.

We find no evidence that this is the case. 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 insignificant and varies in sign across countries suggests that being a younger sibling is not associated with systematically higher levels of human capital. 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).

5. Conclusion

Although economists have studied various aspects of the mass migration from Europe to the US during the 19th century, we know surprisingly little about the economic returns expected by these migrants, and whether migrants were positively or negatively selected from the European population. We use two alternative approaches to estimate the returns to migration in the presence of selection into migration. In the first approach, we use a unique data set of individuals who moved to the US and their brothers who stayed behind to measure the returns to migration from Norway to the US. The premise of this approach is we can eliminate selection across households by comparing brothers who share a common family environment. Our second approach uses a man's rank in his household's birth order as an instrument for migration. We find that, conditional on family size and age, younger brothers in rural areas were more likely to migrate to the US, perhaps because their older siblings were more likely to inherit land under Norway's system of primogeniture.

Our estimates of the returns to migration from Norway to the US range from 52 percent for men leaving urban areas to 120 percent for men leaving rural areas. The returns to migration in the late nineteenth century appear to be substantially smaller than the 200-400 percent returns 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, entry likely kept returns low (in a relative sense).

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 the presence of high inter-generational occupational mobility in the United States in the late nineteenth century (Ferrie, 2004, 2007). For example, sons of unskilled laborers were four times more likely to enter a white collar occupation in the United States than in the United Kingdom in 1880. 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.

In future work, we aim to test how these migrants and their children performed in the US labor markets relative to natives. We will also explore the impact of this negatively-selected migration flow on both the sending and receiving economies. Was the US economy hindered by the arrival of low-skilled migrants or, due to the high rates of upward mobility, was the US able to absorb these migrants without large economic cost? Did Norway benefit from the ability to send its lowest skilled residents to the United States? In addition, we also hope to extend these methods to other source areas, including the United Kingdom, a higher-wage sending country, to determine whether the patterns of selection documented here are broadly generalizable to other European countries.

Data Appendix

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.

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*, forthcoming.

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.

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.

Butcher, Kristin and Anne Casey, "The Effect of Sibling Composition on Women's Education and Earnings", *Quarterly Journal of Economics,* 109 (1994), 531-563

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.

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, "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.

Ferrie, Joseph. "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. "The Path to Convergence: Intergenerational Occupational Mobility In Britain and the U.S. in Three Eras." *Economic Journal*, 117 (2007), C61-C71.

Garg, Ashish and Jonathan Morduch. "Sibling rivalry and the gender gap: Evidence from child health outcomes in Ghana," *Journal of Population Economics*, 11 (1998), 471-493.

"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.

Gould, Eric and Omer Moav, "When is 'Too Much' Inequality Not Enough? The Selection of Israeli Emigrants", manuscript, 2008.

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.

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.

Hatton, Timothy and Jeffrey Williamson, *The Age of Mass Migration: Causes and Economic Impact*. (Oxford: Oxford University Press, 1998).

Hatton, Timothy and Jeffrey Williamson, "International Migration in the Long-Run: Positive Selection, Negative Selection and Policy Williamson and Hatton," NBER Working Paper no. 10529, 2004.

Margo Robert, *Race and Schooling in the South, 1880-1950: An Economic History*. (Chicago: The University of Chicago Press, 1990).

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]

Pande, Rohini, "Selective Gender Differences in Childhood Nutrition and Immunization in Rural India: The Role of Siblings," *Demography*, 40 (2003), 395-418.

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

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.)

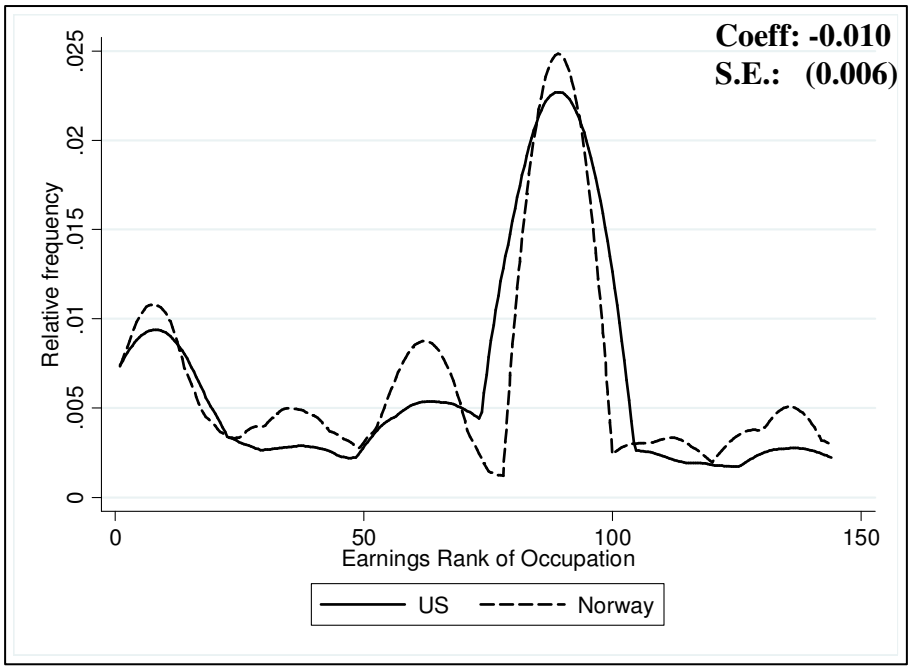
Stock, James H. and Motohiro Yogo. "Testing for Weak Instruments in Linear IV Regression." in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, Donald W. K. Andrews and James H. Stock, eds. (Cambridge: Cambridge University Press, 2005).

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

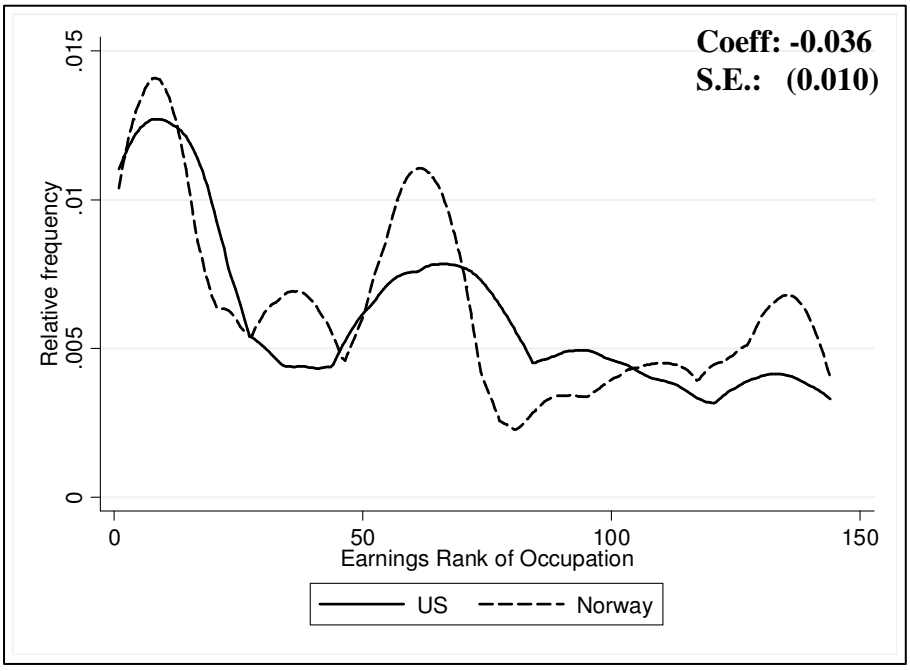
Wegge, Simone A., "Occupational Self-selection of Nineteenth-Century German Emigrants: Evidence from the Principality of Hesse-Cassel," *European Review of Economic History*, 6 (2002), 365-394.

Figure 1: Comparing the occupational distributions of Norwegian-born men in the US and Norway in 1900

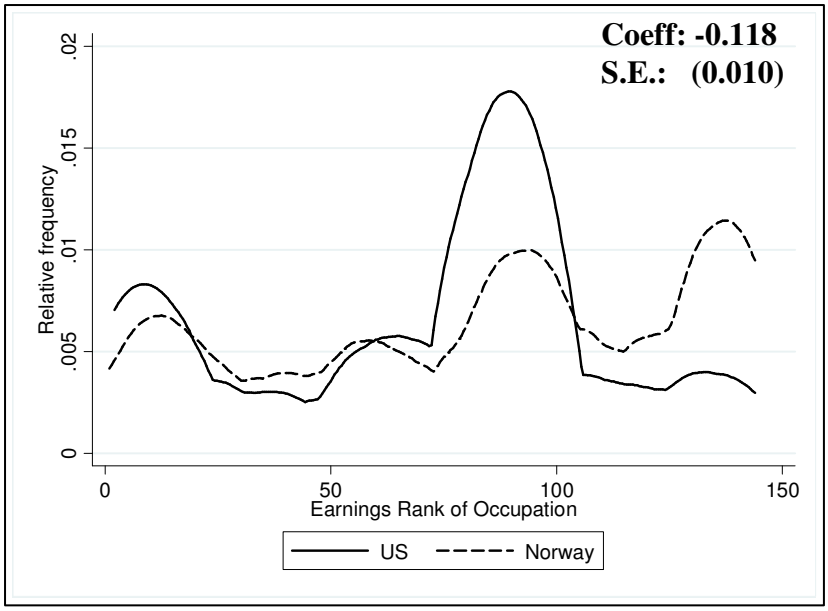
A. <u>Born in rural areas, including farmers</u>



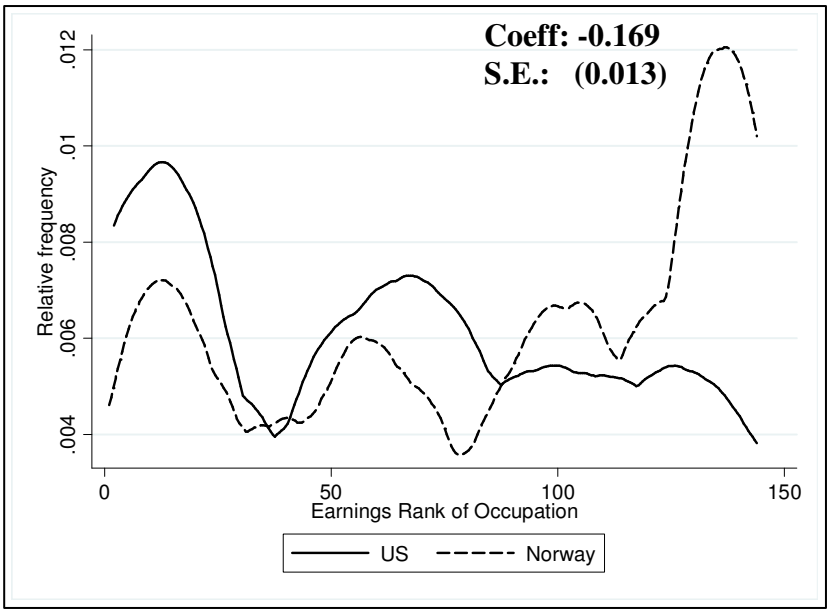
B. <u>Born in rural areas, excluding farmers</u>



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)
B. 1900					
=1 if migrant	n/a	0.146	0.121	n/a	-0.025
		(0.353)	(0.326)		(0.004)
ln(earnings) if live in Norway in 1900	5.675	5.794	5.851	0.118	0.0573
	(0.439)	(0.455)	(0.485)	(0.003)	(0.005)
N	245,765	24,853	10,758	269,618	35,611

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. Columns 4 and 5 present coefficients and standard errors from regressions of the individual characteristics on a dummy variable equal to one for members of the first (second) matched sample. Column 4 looks at 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
ln(earnings)	6.384	6.421	6.433	0.037	0.022
	(0.322)	(0.309)	(0.300)	(0.013)	(0.010)
In 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. Columns 4 and 5 present coefficients and standard errors from regressions of the individual characteristics on a dummy variable equal to one for members of the first (second) matched sample. Column 4 looks at differences between the first matched sample and the total Norwegian population and column 5 looks at differences between the two matched samples. 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 returns 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:
OLS and within-household estimates of the returns 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 dataset. 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 5:
IV estimates of the returns to migration from Norway to the US
Using sibling rank as instrument for migration

	(1)	(2)	(3)
A. First stage			
Sibling rank	0.028	0.018	[Dummies]
	(0.005)	(0.007)	
=1 if owns land		-0.063	-0.063
		(0.018)	(0.018)
Rank * own		0.016	0.016
		(0.007)	(0.007)
*F-stat on rank, rank * own*	*27.42*	*16.42*	*6.09*
B. Second stage			
=1 if migrant	0.802	0.839	0.787
	(0.221)	(0.206)	(0.197)

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 = 7006). 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. 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 6:
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 considers men who lived in the United States in 1900.