Measuring selectivity and returns in the age of mass migration*

[Preliminary and incomplete]

Ran Abramitzky | **Leah Platt Boustan** | **Katherine Eriksson**

Stanford University | UCLA and NBER | UCLA

Abstract

Between 1850 and 1913, the United States absorbed nearly 30 million immigrants from Europe. This paper estimates both the economic return to migrating from Northern Europe to the United States and the degree of selection of the migrant flow. We construct a novel data set of Norway-to-US migrants and their brothers. Because brothers share a common family environment and a portion of their genetic material, the earnings of brothers who remained in Norway provide our best estimate for what migrants' earnings *would have been* had they not migrated. By migrating from Norway to the US, we estimate that Norwegian men could increase their earnings by between 60 and 100 percent. By comparing within-brother estimates to estimates that use both within- and between-household variation, we find evidence for positive selection from rural areas and negative selection from urban areas. An instrumental variables procedure, which uses birth order as an instrument for migration, reinforces the finding of negative selection among migrants from Norwegian cities.

Key words: migration; selection; returns

* Comments welcome at lboustan@econ.ucla.edu. We benefited from conversations with Dora Costa, Avner Greif, Naomi Lamoreaux, Robert Margo, Joel Mokyr, Paul Rhode, Izi Sin and Gavin Wright and members of the KALER group at UCLA. We appreciate Matthew Baird's help with data collection from Ancestry.com. We acknowledge financial support from the California Center for Population Research and the Center for Economic History at UCLA.

1. Introduction

Between 1850 and 1913, an era known as the Age of Mass Migration, the United States absorbed 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.

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 least-able 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 migrant earnings may reflect the fact that migrants were 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, if migrants were negatively selected, the 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 earnings in the US with the earnings of their brothers who remained in Norway. 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 when 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.

Our dataset is based on three complete digitized Censuses, one for the population of the United States in 1880 and two for Norway in 1865 and 1900. The data have been assembled by the North Atlantic Population Project (NAPP). We match Norwegian-born men from the 1880 US Census to their birth families in the 1865 Norwegian data. We then follow all successfully matched men and their brothers forward to either the United States or Norway in 1900. The US forward match is facilitated by the genealogy website Ancestry.com, from which we collect all cases of Norwegian-born men from the 1900 US Census.

In an OLS regression using the complete 1900 Norwegian population and a sample of Norwegian-born men in the US, we estimate a 65 percent return to migration. When comparing within brothers, we find a smaller earnings premium associated with migration of 61.5 percent, with substantially higher returns to migrating from rural areas (65.7 percent) than from urban areas (51.4 percent). In the rural sample, we find that the return to migration estimated within brother pairs is smaller than the return estimated from the whole population, an indication of positive selection across households. For the urban sample, however, we find the opposite, an indication of negative selection.

If brothers share all relevant personal characteristics, our within-household estimates will reveal the "true" return to migration. However, brothers vary in ability. Families with limited resources may have sent the most able or entrepreneurial son to the United States. Alternatively, families may have decided to keep their most able son at home to manage the family business or take care of the parents in old age.

Such concerns encourage us to develop a complementary estimation strategy. Specifically, we compare our migrant sample with Norwegian-born men in the same age range, using rank in household birth order as an instrument for migration. Migration decisions were often made by family networks. We find that, conditional on age and family size, younger siblings were less likely to migrate, perhaps because older brothers were sent to the US to establish a homestead before sending for other family members.[1] Older brothers may also have been more likely to receive financial transfers to defray migration costs. The identifying assumption of this IV approach is that being an older brother, while increasing the probability of migration, does not directly affect earnings. We provide supporting evidence for this assumption, demonstrating that there is little correlation between a man's rank among his siblings and his earnings among men who remained in Norway or Norwegian men who lived in the United States.

Our first stage regressions confirm that being a younger sibling had a large negative effect on the likelihood of moving to the United States. Our second-stage regressions of earnings on the instrumented dummy for living in the US imply returns to migration of between 80 and 100 percent. As in the within-household analysis, the IV estimates of the returns to moving from

[1] Even if some fraction of the younger brothers in our sample eventually join their older brother in the US, we will still expect to find a positive relationship between birth order and the probability of migration.

urban areas are substantially larger than the OLS estimates, corroborating our finding of negative selection among migrants from Norwegian cities.

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 Norway-to-US migrants to their brothers who remained in Norway. Section 4 discusses the results from our two estimation strategies – namely conducting a within-household analysis and using household gender composition 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).[2] 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. The Norway-to-US migration is thus quantitatively important *per se*.

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

[2] This section is based on Hatton and Williamson (1998) and Chiswick (2003).

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 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 75 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 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 by using information on the family background of prospective migrants to predict migrant status.

3. Data and Matching

A. Data

Our goal is to identify pairs of Norwegian brothers containing at least one Norway-to-US migrant and observe the labor market outcomes of both brothers later in life. For this procedure, we rely on three complete digitized Censuses, one in the US in 1880 and two in Norway in 1865 and 1900. These datasets are archived at the North Atlantic Population Project (NAPP). We match Norwegian-born men from the 1880 US Census to their birth families in the 1865 Norwegian data. We then follow all successfully matched men and their brothers forward to either the United States or Norway in 1900. The US forward match is facilitated by the genealogy website Ancestry.com, from which we collect all cases of Norwegian-born men from the 1900 US Census records. 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 income. Instead, we assign men the mean income earned by members of their occupation.[3] We collect the income information from the 1901 Cost of Living Survey in the United States (Haines and Preston, 1991) and from tabulations of mean income by occupation for the year 1900 published by Statistics Norway and other sources (*Statistik Aarbog*, 1900; Grytten, 2007).[4] The Cost of Living Survey reports income information for more than 300 occupations. At least one member of our sample is employed in 138 of these categories. Table 1 reports the ten most common occupations for our sample of matched brothers in Norway and the United States. Half of each group worked in farm occupations. Other common occupations include carpenters, merchants and sawmill operatives. Migrants to the US were more likely to report being general laborers.

These sources do not report information on earnings for a few large occupations, including farmers in both countries, 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 to 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 the data sources and assumptions underlying these estimates. With the exception

[3] For men living in the US, we code occupation by hand using the digital images of Census manuscripts available on Ancestry.com.
[4] We convert Norwegian wages to PPP-adjusted US dollars using the 1900 exchange rate and price levels from Global Financial Data.

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 80th or 90th 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 Procedure

We construct our dataset by locating Norwegian migrants in their birth families in 1865 and tracking the resulting sample of migrants and their brothers in the labor market. Our baseline method ("**Match 1**") uses an iterative matching strategy pioneered by Ferrie (1996). We describe this procedure in detail:

(1) There are 100,937 Norwegian-born men in the 1880 US Census. 37,805 of them are between the ages of 15-30 in 1880 and thus were likely to be living with their birth families in 1865.

(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 25,297 of these observations that are unique within their first name, last name and birth year.[5]

(3) We match the 25,297 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).[6] Steps 3 and 4 produce a sample of 3,991 migrants and 6,130 brothers.

(5) We search for migrants and their brothers in 1900. We attempt to match migrants in the US first, looking only for unmatched migrants in Norway (*vice versa* for

[5] Men with uncommon first or last names have a higher probability of being included in our sample. The average migrant in our sample has a name that occurs 20 times in the 1900 Norwegian Census, compared to population average of 150. Migrants' brothers fall somewhere in between (average = 96). We find that the commonness of a man's name is strongly correlated with his labor market earnings in 1900; the average man with a unique name earns 12 percent more than the average man whose name is shared by 100 other Norwegians. However, the commonness of one's name is not associated with earnings within families.

[6] 97 percent of our pairs are sons of the household head. Grandsons of the household head may be cousins, rather than brothers.

brothers). All US matches are conducted using the iterative procedure described in step 3. Norwegian matches benefit from additional information about municipality and province of birth. We conduct Norwegian matches first by name, birth year, and municipality of birth.[7] For all observations left unmatched by this procedure, we also conduct a match by name, birth year and province of birth.

This procedure creates a sample of 2,994 observations – 1,429 men who lived in the United States in 1900 and 1,592 who lived in Norway. Step 3 achieves a backwards match rate of 11 percent among all Norwegian-born men or 16 percent among men with a unique name-birth year combination in the United States in 1880. These rates are comparable to Ferrie's (1996) backwards match rates within the United States of nine and 19 percent respectively. In our forward match, we find 48 percent of our eligible migrants and 22 percent of their brothers. The migrant match rate is higher than the brother match rate because migrants are already restricted to be unique by name and age.

We are concerned that the iterative nature of this method will produce false matches. Case in point: we match 32 percent of eligible migrants to the United States in 1900 and 16 percent to Norway. The implied return migration rate (35 percent) is higher than the 25 percent return migration rate calculated by the Norwegian Bureau of Statistics for 1880 to 1900 (Semmingsen, 1978, p. 120).[8] False matches may occur because we stop searching for matches if we find an exact match in the first step of our iterative process. Thus, we may keep men in the sample who have both an exact match <u>and</u> a "close" match (within a one- or two-year band around the reported birth year). As a result, in our final step, we have the potential to identify

[7] Men in this cohort reported 751 municipalities of birth in 1865 and 525 municipalities of birth in 1900. We suspect that villages consolidated into larger towns over time and that men reported the current, rather than the historical, name of their birth place in 1900.

[8] 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.

multiple matches in 1900. Given that we search in the United States and Norway sequentially, some of these multiple matches will be averted. However, consider a case in which the "real" individual lived in the United States in 1900 but misreported his age.[9] Because the match cannot be successfully completed in the US, we risk identifying a false match in Norway instead.

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 and *three-year age band* in both locations. 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 sample of 1,313 observations – 677 men who lived in the United States in 1900 and 799 who lived in Norway. We find 29 percent of the initial migrants in the US in 1900 and 10.8 percent in Norway (an implied return migration rate of 27 percent).

Table 2 compares attributes of the two matched samples. 47 and 49 percent of the men in Match 1 and Match 2, respectively, migrated to the United States; this difference is not statistically significant. Note that because our samples are designed to contain at least one migrant per family, the lifetime migration rates are higher than the 34 percent rate in the general Norwegian population.[10] In both samples, the average man was 44 years old in 1900. In 1865, at the average age of nine, he lived with 3.5 siblings, 44 percent of whom were sisters.[11] The one quantitatively (and statistically) important difference between the two samples is that men in Match 2 were 12 percentage points more likely to have been born in an urban area, perhaps because rural families used a narrower array of given names (Gjerde, 1985, p. 48).

---

[9] Age heaping around certain focal ages (for example, 40 or 50) is a particular concern in an era before birth registration and for a match conducted over a 35 year time horizon.

[10] Over 700,000 Norwegians left the country from 1866 to 1915 on a base population (in 1900) of 2.2 million (Semmingsen ,1978, p. 99).

[11] 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.

4. Estimating the return to migration

A. A naïve approach: Comparing earnings of Norwegian-born men in the US and Norway

Our data set of Norwegian brothers provides a unique opportunity to address selection bias in estimating the returns to migration. For a benchmark, however, we begin by comparing the earnings of all Norwegian-born men living in the United States to all men who remained in Norway in 1900. While this approach does not account for selection, it is commonly used to estimate the return to migration when Census data are available for both sending and receiving country (Hanson, 2006).

We combine working-age men (15-70 years old) in the 100 percent Norwegian Census in 1900 with comparable Norwegian-born men in the 1 percent sample of the 1900 US Census from the Integrated Public-Use Microdata Series (IPUMS).[12] 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.

The coefficient of interest 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 65 percent more than men who remained in Norway in 1900. In the next columns, we augment equation 1 with interactions between migration status and place of residence and age. The earnings gap between residents of Norway and the United States is larger in rural areas (68 percent) than in urban areas (60 percent). The returns to migration increase with both age and years spent in the United States. The average 15 year old migrant earned 44

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

percent more than the average 15 year old worker in Norway, whereas the average 45 year old migrant earned 71 percent more than the average 45 year old worker in Norway.[13]

B. Within-household analysis

We hesitate to interpret the 65 percent earnings gap between the United States and Norway as the return to migration. In particular, migration status is likely to be correlated with unobserved individual characteristics – such as ability, motivation, or willingness to take risks – that are captured in the error term (ε_i). If ability, for instance, is positively correlated with both an individual's earnings and his propensity to migrate, our estimate of β_1 will be biased upward.

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 uncorrelated across individuals.[14] 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 between-occupation selection that we will call θ.[15] Adding household fixed effects to equation 1 absorbs the common portion of the error (α_j), which, under the restrictive assumptions above, will also eliminate θ. If households that send migrants to the United States are positively selected

---

[13] The pattern of returns by age could also reflect a differential return migration rate among the unsuccessful.
[14] We find strong empirical evidence that brothers are more similar to each other than they are to the rest of the population. Using the complete 1865 and 1900 Censuses, we document that the earnings of brothers who both remain in Norway have a correlation coefficient of 0.180, whereas the earnings of two randomly selected men is -0.005.
[15] If we think about the component of the error term that is shared between siblings as a single omitted variable (for example, ability), θ is equal to the coefficient on ability from a regression of log earnings on migration status and ability multiplied by the coefficient from a regression of migration status on ability.

on ability, then we would expect the estimate derived from comparing brothers in the same household to be less than the OLS coefficient ($\beta^H < \beta_1$).[16]

If the individual component of the error term (ν_{ij}) is uncorrelated with migration status, our within-household estimate will reveal the "true" return to migration. However, 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 but will still be lower than the "naïve" estimate ($\beta^{true} < \beta^H < \beta_1$).

Table 4 estimates OLS regressions with and without household fixed effects for our baseline sample (Match 1) and our robustness sample (Match 2).[17] The regressions without household fixed effects can be compared to the estimates in Table 3. The matched samples and the full population produce similar estimates; compare a 65 percent return in the population with 63 and 68 percent in the matched samples.[18]

More importantly, the matched samples allow us to compare estimates of the return to migration that rely only on variation within households to those that use both within- and between-household variation. We find that adding household fixed effects reduces the estimated return to migration in both samples. In Match 1, the within-household estimate of the return to migration is 61.5 percent (compared with 68.4 percent), while in Match 2, the within-household estimate of the return to migration is 53.8 percent (compared to 63.0 percent). This pattern

[16] See 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.

[17] Around 40 percent of the migrants in both matched samples have a brother who can also be located in 1900. This subset identifies β^H. Regressing earnings on migration status without household fixed effects in this sub-sample produces nearly identical coefficients to those reported in the first row of Table 4.

[18] Because of our sample selection process, our matched samples contain men between the ages of 35 and 50.

suggests that high-ability families were more likely to send migrants to the United States. Comparing the coefficients suggests that 10-15 percent of the naïve "return to migration" may instead be measuring positive selection across families. Of course, positive selection *within families* may explain a further portion of the estimated return and is not captured here.

In the second and third columns of Table 4, we subdivide the sample by place of birth. More than two-thirds of both matched samples lived in rural areas in 1865. The return to migration was higher for men born in rural areas than for their counterparts born in urban areas (65.7 percent versus 51.4 percent).[19] The lower return among urban men may partially reflect the lack of data for the wages and salaries of white collar workers in Norway.

Surprisingly, we only find evidence of positive selection among rural families. In contrast, among urban families, we estimate a larger return to migration in the within-household specification. Thus the migration flow from Norwegian cities and towns appears to have been drawn from households with lower ability, fewer connections, or less wealth. Such differences in the selection process would be expected 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, estimating 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.

While it seems plausible that brothers are more similar than two randomly selected members of the population, it is important to note that relying on differences in migration activity between brother pairs might accentuate the effect of measurement error on earnings,

[19] We find similar results for regressions that exclude farmers and fisherman, the two main categories of workers whose wages we needed to estimate.

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

resulting in an attenuation bias of the within-household estimates. (Griliches, 1979; Ashenfelter and Krueger, 1994). Intuitively, this attenuation bias arises because narrowing the comparison to brothers reduces the variation in the independent variable (migration status) without reducing measurement error. In particular, the smaller within-household estimates in the rural sample – which we have attributed to positive selection – may simply be driven by measurement error, while the degree of negative selection identified in urban areas may be larger than it appears in this context. The next section will provide results from an instrumental variables analysis that both corrects for any original measurement error in the OLS estimates. Comparing these two approaches will allow us to put bounds on the share of the estimated return due to selection.

C. Instrumental variable analysis

The difficulty of interpreting within-household estimates in the presence of measurement error and 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 or household characteristic that is correlated with the propensity to migrate but is not otherwise associated with labor market potential.

Passage to the United States was often funded by family networks and, in return, the migrant was expected to remit a portion of his earnings to his relatives in Norway and/or to facilitate the migration of other family members to the US (Hatton and Williamson, 1998). As a result, we expect that a man's placement in the birth order will influence his likelihood of being sent to 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 therefore will have less to gain from migration. 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, during this episode of mass migration, families may expect that all or most of their members will eventually relocate to the US. In this case, older brothers may migrate first in order to lay the groundwork for the arrival of other family members.

Our argument implies that older siblings may be more likely to migrate to the United States than younger siblings due to the happenstance of their rank in the birth order (or *vice versa*). Of course, in addition to migration assistance, birth order may affect a whole stream of human capital investments in nutrition, education and occupational training. Thus, sibling rank may not be excludable from the earnings equation. We test for this possibility below and find no evidence that birth order affects the earnings of men who remained in Norway or the earnings of Norwegian migrants in the United States. This 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).

For the purpose of using birth order among a man's siblings as an instrumental variable for migration, we broaden our comparison beyond members of a migrant's own household to all Norwegian-born men in the relevant age range (0-15 in 1865) who were unique within their name and age in 1865 and, thus, had the potential to match with a Norway-to-US migrant. Our first stage equation relates an individual's migration status to his rank among his siblings observed in his childhood household in 1865. We regress:

$$\text{Migrant}_{ij} = \alpha + \gamma_1(\text{sibling rank}_{ij}) + \Gamma A_{ij} + \Psi F_j + \Omega P_j + \varepsilon_{ij} \qquad (2)$$

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, total number of siblings in the household, and province of residence in 1865. γ_1 can be interpreted as comparing two men with the same age and the same number of siblings who differ only in their placement in the birth order.

Because we only observe the childhood household at a single point in time, we will disproportionately mis-measure the rank of older men among their siblings. The oldest men in our sample (15 in 1865) are likely to have unobserved older siblings who already left home and thus may be misclassified as oldest siblings. For example, given the empirical relationship between number of siblings and age, we predict that 15 year olds should have 3.6 siblings – yet we observe the typical 15 year old living with only 3.2 of his siblings. In contrast, even if we understate the number of younger siblings that the typical one or two year old will eventually have, we will not mis-measure his rank among siblings. As a result, we restrict our IV analysis to the younger half of our sample (0-8 years old in 1865), but our findings are robust to using alternate cut-off ages.[21]

We report estimates of γ_1 in the first row of Table 5. Being a younger brother decreases the probability of migration, with each increase in sibling rank reducing the likelihood of migration by 1-2 percentage points (relative to the sample mean of 47 percent).[22] Our second-stage regressions re-estimate equation 1 using sibling rank as an instrument for migration.[23] For comparison, we reproduce the OLS estimates in the second row of Table 5 and report the IV

[21] We also drop the 0.65 percent of the sample in households for which the spouse of the household head is reported to be less than 20 years old. We suspect that many of these "spouses" are actually misclassified daughters of the household head. Men in these households have a sibling rank that is nearly 0.5 below the rest of the sample, a value which is consistent with the presence of a missing sibling in the birth order (given that the mis-classified daughter will be older than the sample individual half of the time).

[22] The F-statistic is greater than ten in all cases, which minimizes concerns about weak instruments (Stock and Yogo, 2005).

[23] The second stage regressions also include dummy variables for single year of age, total number of siblings in the household and province of residence in 1865.

estimates in the third. We estimate an overall return to migration of 107 percent with a lower – but more precisely estimated – return of 83.4 percent for men born in rural areas.

Unlike the within-household analysis, we find here that correcting for the underlying selection process heightens the estimated return to migration ($\beta^{IV} > \beta_1$), suggesting that, if anything, migrants appear to be negatively selected. That is, leaving aside questions of measurement error for the moment, it appears that men whose migration decisions differ only due to their rank in the household birth order experience higher returns than are found in the sample at large. This difference implies that the return to migration would have been 82 percent higher than the value we found in the naïve (OLS) regression if not for this negative selection.

How can we reconcile the apparent negative selection here with the within-household estimates? Columns 2 and 3 subdivide the sample into men born in rural and urban areas. Comparing these two columns reveals that, consistent with our earlier findings, the overall pattern of negative selection is primarily driven by the urban sample. Indeed, the gap in the naïve migration returns estimated for urban and rural men is entirely explained by negative selection in urban areas. The return to migration experienced by men moving from urban areas (44 percent) appears to be artificially dampened by negative selection; when we instrument for the migration decision, we instead find that urban men achieve a return of 105 percent.

In both empirical approaches, we find that controlling for selection heightens the return to migration from urban areas, a pattern consistent with negative selection. Determining the extent of this selection, however, is complicated by the presence of measurement error. The within-household estimates imply that the estimated return to migration is only 11 percent too low due to negative selection, but these estimates may suffer from attenuation bias. In contrast, the IV estimates suggest that the estimated return to migration is over 100 percent too small, but the IV estimates may be larger than OLS simply because they correct for measurement error. For the rural sample, our conclusions must be even more circumspect. The within-household estimates present evidence of slight positive selection (explaining 18 percent of the naïve return) but the IV estimates point toward negative selection (suggesting that the naïve return is 21 percent too small). However, we cannot rule out that the within-household estimates are smaller than OLS due to attenuation, while the IV estimates are larger than OLS because of corrections for measurement error. In this case, we can only conclude that the selection from rural areas is mild and is centered around zero.

The validity of this instrumental variables procedure rests on the assumption that one's 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 an older 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. While being first rather than second born is associated with small positive increases in earnings in Norway (0.4 percent for men born in rural areas and 1.5 percent for men born in urban areas), these estimates are statistically indistinguishable from zero. Even were these effects to be statistically significant, they are too small to explain the doubling in the estimated returns to migration in the IV specification for men born in urban areas.

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 while accounting for migrant selection. 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 that selection is less likely to be important when comparing brothers who share a common family environment and a portion of their genetic material. Thus, comparing the earnings of brothers who stayed with the earnings of brothers who moved allows us to get a more precise estimate of the true returns to migration.

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, older brothers were more likely to migrate. Because migration decisions were made in a family context, older brothers may have been encouraged to migrate in order to establish a household and facilitate the migration of other family members. This exogenous increase in the probability of migration allows us to identify the true returns to migration.

Our estimates of the returns to migration from Norway to the US range from 60 to 100 percent. This value is a somewhat higher than the return to spending four years in high school, which Goldin and Katz (2000) estimate provided a return of 40 percent. Furthermore, we find that selection can generate substantial bias in the naïve estimates of the return to migration. For men born in urban areas, we find that a naïve estimate of the return to migration is at least 11 percent – and may be up to 140 percent – too low due to negative selection. In contrast, any selection among men born in rural areas appears to have been mild. The US economy may have been hindered by the arrival of negatively selected urban migrants. Understanding the different patterns of selection in rural and urban areas is a fruitful direction for future research.

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* (1998).

Abramitzky, Ran and Fabio Braggion. "Migration and Human Capital: Self Selection of Indentured Servants to the Americas," *Journal of Economic History* (2006), 66:4, 882-905.

Abramitzky, Ran. "The Effect of Redistribution on Migration: Evidence from the Israeli Kibbutz", *Journal of Public Economics* (2008), Forthcoming.

Altonji, Joseph P. and Thomas A. Dunn. "Using Siblings to Estimate the Effect of School Quality on Wages," *Review of Economics and Statistics* (1996).

Ashenfelter, Orley and Alan Krueger. "Estimates of the Economic Return to Schooling from a New Sample of Twins," *American Economic Review* (1994).

Atack, Jeremy and Fred Bateman. "'Matchmaker, Matchmaker, Make Me a Match': A General Personal Computer-Based Matching Program for Historical Research," *Historical Methods* (1992) 25: 53-65.

Behrman, Jere R., Mark R. Rosenzweig, and Paul Taubman. "College Choice and Wages: Estimates Using Data on Female Twins." *Review of Economics and Statistics* 78:4 (1996), 672-685.

Behrman, Jere R. and Mark R. Rosenzweig, "Does Increasing Women's Schooling Raise the Schooling of the Next Generation?" American Economic Review 92:1 (2002), 323-334.

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

------- "Immigration and Self-Selection," in *Immigration, Trade and the Labor Market*, ed. by J. Abowd, and R. Freeman. The University of Chicago Press, Chicago (1991).

------- "The Economics of Immigration," *Journal of Economic Literature* (1994), 32(4), 1667-1717.

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

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.

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* (2005), 113(2), 239-281.

Chiswick Barry. "Are Immigrants Favorably Self-Selected?," *American Economic Review* (1999), 89(2), 181-185.

------- "Are Immigrants Favorably Self-Selected? An Economic Analysis," in *Migration Theory: Talking Across Disciplines* (2000), ed. by C. D. Brettell, and J. F. Hollifield, pp. 61-76. Routledge, New York.

Chiswick, Barry and Timothy Hatton, "International Migration and the Integration of the Labor Markets," in *Globalization in Historical Perspective* (2003).

Feliciano, Cynthia. "Educational Selectivity in US Immigration: How Do Immigrants Compare to Those Left Behind?" *Demography*, 42:1 (2005), p. 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* (1996).

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," (1999) Working paper.

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

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

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

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

Goldin, Claudia and Lawrence F. Katz. "Education and Income in the Early Twentieth Century: Evidence from the Prairies." *Journal of Economic History* (2000).

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(5), (1979), p. S37-S64.

Grytten, Ola H. "A Consumer Price Index for Norway, 1516-2003." In *Historical Monetary Statistics for Norway, 1819-2003.* Eds., Øyvind Eitrheim, Jan T. Klovland and Jan F. Qvigstad. 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*. Eds., Øyvind Eitrheim, Jan T. Klovland and Jan F. Qvigstad. 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*, December 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*, 20:3 (1994), p. 533-559.

Hatton, Timothy and Jeffrey Williamson, *The Age of Mass Migration: Causes and Economic Impact*, 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*, 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]

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* (2004).

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

Statistiske Centralbyraa. *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." D.W.K. Andrews and J.H. Stock, *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. Cambridge: Cambridge University Press, 2005, pp. 80-108.

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 (3): 365-394, 2002.

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

Rank	Occupation	Frequency	Percentage	Earnings
1	Farmers and Planters	649	45.4	$691
2	Laborers (General)	79	5.53	373
3	Farm Laborers	58	4.06	255
4	Carpenters and Joiners	44	3.08	630
5	Painters, Glaziers, and Varnishers	37	2.89	624
6	Merchants – Not Specified	23	1.61	1100
7	Saw and Planing Mill Workers	18	1.26	572
8	Draymen, Teamsters, and Expressmen	16	1.12	546
9	Sailors	15	1.05	467
10	Merchants – General Store	15	1.05	1100
11	Salesmen and Saleswomen	15	1.05	680
Total		969	67.8	

A. Top 10 occupations in matched sample, US 1900

N= 1429. Census occupation categories. 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, Norway 1900

Rank	Occupation	Frequency	Percentage	Earnings
1	General Farmers	497	31.2	393
2	Husbandmen or Cottars	105	6.60	113
3	Farmer and Fisherman	85	5.34	321
4	Farm Workers	82	5.15	175
5	Carpenters	64	4.02	312
6	Dealer, Merchant, etc.	57	3.58	837
7	Sawyers and Sawmill Operatives	49	3.08	269
8	Boot and Shoe Makers and Repairers	40	2.51	276
9	Fishermen	39	2.45	248
10	Lumbermen and Loggers	35	2.20	256
Total		1053	66.1	

N= 1592. 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 is estimated using data from the Norwegian Census of Agriculture. The Data Appendix provides details on this procedure.

Table 2:
Comparing the two matched samples in 1900

	Match 1	Match 2	Difference	p-value
Age	43.98	44.16	-0.180	0.358
	(5.90)	(5.78)		
=1 if migrant	0.468	0.494	-0.025	0.127
	(0.499)	(0.500)		
=1 if urban, 1865	0.223	0.344	-0.121	**0.000**
	(0.417)	(0.475)		
Number siblings	3.56	3.66	-0.100	0.149
	(1.93)	(1.99)		
=1 if oldest sib	0.444	0.433	0.011	0.489
	(0.497)	(0.496)		
Share sisters	0.448	0.442	0.004	0.617
	(0.313)	(0.312)		

Notes: Columns 1 and 2 contains means and standard deviations (in parentheses) of individual characteristics 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 three-year age band. Column 3 presents the difference in means between the two samples and column 4 contains the p-value for the null hypothesis that this difference is equal to zero. The number of siblings variable is exclusive of the individual himself.

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

Dependent variable = ln(earnings)

	(1)	(2)	(3)	(4)	(5)
=1 if migrant	0.665 (0.007)	0.444 (0.025)	0.681 (0.009)	0.552 (0.015)	0.450 (0.025)
Age		0.025 (0.0001)			0.025 (0.0001)
Age · migrant		0.006 (0.001)			0.004 (0.001)
=1 if urban			0.197 (0.001)		
Urb · migrant			-0.081 (0.016)		
Years in US				0.006 (0.001)	0.003 (0.001)

Source: 100 percent 1900 Norwegian Census and 1 percent 1900 US Census sample (IPUMS).
Notes: *N*= 566,002. Men aged 15-70. All regressions include age and age squared. The urban variable refers to place of residence in 1900.

Table 4:
Within-household estimates of the returns to migration from Norway to the US

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

	All	Rural, 1865	Urban, 1865
Panel A: Match 1			
OLS	0.684	0.720	0.464
	(0.014)	(0.016)	(0.033)
OLS - Within household	0.615	0.657	0.514
	(0.030)	(0.034)	(0.069)
N	2972	2303	669
Panel B: Match 2			
OLS	0.630	0.705	0.417
	(0.023)	(0.029)	(0.039)
OLS - Within household	0.538	0.581	0.469
	(0.046)	(0.059)	(0.073)
N	1269	817	452

Notes**:** Each cell contains coefficient estimates of log earnings on a dummy variable equal to one for individuals living in the United States in 1900. Standard errors are reported in parentheses. 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 three-year age band. In each panel, the first row presents coefficients from an OLS regression without household fixed effects (comparable to Table 3). 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

	All	Rural	Urban
First stage	-0.017	-0.015	-0.021
	(0.003)	(0.003)	(0.007)
F-statistic	41.67	28.77	10.33
OLS	0.625	0.689	0.445
	(0.017)	(0.020)	(0.036)
IV	1.074	0.834	1.057
	(0.314)	(0.372)	(0.605)
N	12116	9465	2651

Notes: Results for men in the first matched sample (Match 1). The first row in each panel presents the coefficient of interest from first stage regressions of migration status on sibling rank (1= oldest) and a series of dummy variables for age, number of siblings and province of residence. The second and third rows contain OLS and 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. Standard errors are reported in parentheses. The second and third columns conduct similar analyses for men who lived in rural or urban areas respectively in 1865.

Table 6:
Testing the exclusion restriction: The relationship between sibling rank and earnings among migrants and stayers

Dependent variable = ln(earnings)

	Rural, 1865	Urban, 1865
US in 1900	-0.0003	0.008
	(0.014)	(0.024)
Norway in 1900	-0.004	-0.015
	(0.006)	(0.013)

Notes: Results for men in the first matched sample (Match 1). Each cell contains a coefficient estimate from the regression of log earnings on sibling rank (1= oldest) and a series of dummy variables for age, number of siblings and province of residence. Standard errors are reported in parentheses. The first row restricts attention to men who lived in the United States in 1900, while the second row considers men who lived in Norway in 1900. The columns conduct similar analyses for men who lived in rural or urban areas respectively in 1865.