Intergenerational Occupational Mobility Across Three Continents

Santiago Pérez

Santiago Pérez is Assistant Professor, Department of Economics, University of California, Davis, 1118 Social Sciences and Humanities, Davis, CA 95616. E-mail: seperez@ucdavis.edu.

An earlier version of this paper was part of my dissertation completed at Stanford University. I thank my advisor Ran Abramitzky, as well as Arun Chandrasekhar, Victor Lavy, Melanie Morten and Gavin Wright for outstanding guidance as part of my dissertation committee. I have also benefitted from feedback from participants at the UC Davis HIA/ISS seminar and seminar participants at the University of Arizona and PUC Chile. I acknowledge the financial support of the Stanford University Economics Department, the Economic History Association, the Stanford Center for International Development and the Leonard W. Ely and Shirley R. Ely Graduate Student Fund Fellowship.

I compare rates of intergenerational occupational mobility across four countries in the late 19th century: 1869-1895 Argentina, 1850-1880 United States, 1851-1881 Britain, and 1865-1900 Norway. Argentina and the US had similar levels of intergenerational mobility, and these levels were above those of Britain and Norway. These findings suggest that the higher mobility of 19th-century US relative to Britain (Long and Ferrie, 2013) might not have been a reflection of "American exceptionalism", but rather a reflection of more widespread differences between settler economies of the New World and Europe.

Throughout the 19th century, several observers pointed to the "exceptional" social mobility experienced by the US population. For instance, when comparing the social class structures in the US and Europe, Marx (1852) argued that in the US social classes "have not yet become fixed but continually change and interchange their elements in constant flux."[1] Recent research indeed shows that the US exhibited higher intergenerational occupational mobility than Britain in the second half of the 19th century (Long and Ferrie, 2013).

In this paper, I ask whether the higher mobility of the US relative to Britain was a reflection of "American exceptionalism" or, rather, a reflection of more widespread differences between areas of recent European settlement and the Old World. To do so, I compare rates of intergenerational occupational mobility across four countries in the late 19th century: 1869-1895 Argentina, 1850-1880 United States, 1851-1881 Britain, and 1865-1900 Norway.

The Argentina-US comparison is of special relevance. Both countries shared many of the characteristics that have been suggested as explanations for the high mobility of 19th-century US: they were areas of recent European settlement, with sparse (compared to other countries in the Americas) indigenous populations and large areas of unexploited land (the "frontier"). Both countries also attracted a large number of European immigrants and experienced rapid growth during the second half of the 19th century. The contrast with Norway is also revealing since, unlike Britain, Norway had a similar occupational structure, characterized by a large fraction of workers in the

[1] Similar remarks were made by De Tocqueville (1835), among others.

primary sector, to Argentina and the US in the mid-19th century.

To conduct the analysis, I constructed data linking fathers and sons across national censuses of population for these four countries. In these datasets, I observe an individual's occupation in adulthood, as well the occupation held by his father when the individual was a child. I use these data to analyze the extent to which a father's occupation provided his sons with an advantage in accessing a given occupation in adulthood. All these datasets were built using similar underlying sources –national censuses of population– and the same linking strategy based on names, year of birth and place of birth, thus facilitating the comparison across countries.

I find that Argentina and the US had similar levels of intergenerational occupational mobility, and that these levels were above those of Britain and Norway. These findings suggest that the higher mobility of 19th-century US relative to Britain might not have been a reflection of "American exceptionalism", but rather a manifestation of more general differences between settler economies of the New World and Europe.

Extending the comparison beyond two countries enables me to make some progress on the question of *why* mobility levels might have been higher in Argentina and the US. I first argue that the relative abundance of land in Argentina and the US made the transition into independent farming relatively easier in Argentina and the US than in Europe. This explanation is also consistent with the fact that rates of intergenerational mobility in the US appeared to have converged to those in Europe during the 20th century, after frontier expansion had concluded (Long and Ferrie, 2007; Modalsli, 2017). My results also suggest that differences in the fraction of immigrants in the population were unlikely to be the main reason for the observed differences, as mobility was higher in Argentina and the US when just focusing on the children of natives.

One concern with using linked data in an international comparison of intergenerational mobility is that the selection into the samples or the prevalence of incorrect matches might be different across countries. In the context of this paper, if the rate of incorrect matches were higher in Argentina and the US than in Britain and Norway, I would mechanically find higher rates of mobility in the former two countries. To address this concern, I show that the results are similar when using a more conservatively linked sample *only* for Argentina and the US, thus biasing the samples against the main finding of the paper –that is, that Argentina and the US had higher mobility than Britain and Norway. I also show that, to rationalize the observed differences in mobility between the two groups of countries, the share of incorrect matches in the Argentine and US samples would have needed to be implausibly higher than in the British and Norwegian data.

I emphasize that these results pertain to *occupational* mobility, as measuring *income* mobility is not feasible for any of these countries in this period. The study of occupational mobility has a long tradition in sociology (see, for instance Jonsson et al. (2011)), and has recently attracted attention in economics (Long and Ferrie, 2013). The extent to which occupation-based measures of mobility agree with income-based measures of mobility is an empirical question. Recent research comparing contemporary levels of intergenerational mobility in the US to those in the early 20th century (Feigenbaum, 2017) finds a similar pattern (higher mobility in the past than contemporarily) regardless of the use of occupation-based or income-based mobility measures. In any case, occupation is the only measure of social status that can be consistently tracked in these four countries during this period.

This paper is closely related to research on comparative intergenerational mobility over the long run. Examples in this literature include Long and Ferrie (2013), Olivetti and Paserman (2015), Modalsli (2017), and Feigenbaum (2017). Unlike my paper, these studies focus exclusively on Europe and the US.[2]

More broadly, the evidence on historical social mobility for contemporary developing countries is limited.[3] Given the limited number of studies on historical mobility in the developing world, *comparative* studies including developing countries are even rarer. One exception is Clark (2014), which uses the status information contained in surnames to characterize social mobility for several countries and time periods, including Chile, China and India.[4]

[2]There is also a large literature on contemporary levels of intergenerational mobility, both within specific countries and in a comparative perspective. Examples in this literature include Chetty et al. (2014a), Chetty et al. (2014b) and Solon (2002), among many others.

[3]Two exceptions are Cilliers, Fourie et al. (2016) and Chen, Naidu, Yu, and Yuchtman (2015). Cilliers, Fourie et al. (2016) studies intergenerational mobility in 19th-century South Africa. This paper uses genealogical data (mostly marriage and baptism records) rather than national censuses of population, thus complicating a direct comparison to the evidence presented here. Chen et al. (2015) study mobility in educational attainment in China over the 20th-century.

[4]The author finds high and similar levels of persistence in social status across countries and periods. This difference in results with respect to Clark (2014) can stem from two main reasons. First, my results pertain to mobility across two generations, while Clark measures mobility over the course of several generations. Second, while I measure mobility across families (or more precisely, across father-son pairs), Clark measures the persistence of elite status across surname groups: The two measures of mobility will differ as long as there is mobility *within* surname groups. See Section III in Solon (2015) for a discussion on why group-level estimates of mobility might differ from individual-level estimates.

Data and Empirical Strategy

Creating the Linked Samples

I applied a consistent methodology to construct father-son linked datasets for Argentina, Britain, Norway and the US in the second half of the 19th century. In each of these datasets, I observe boys who resided with their father in an earlier census, and I then link them to their labor market outcomes in adulthood in a later census. I focus on men because in the US, British and Norwegian samples it is not possible to follow women across censuses due to name changes upon marriage.

To construct the Argentine sample, I followed males through the 1869 and 1895 Argentine censuses, using the indexes available in the genealogy website FamilySearch.org. The 1869 Argentine census was restricted to the 14 provinces that officially constituted the country at the time. The remaining areas of the country were considered national territories, and had an estimated population of approximately 90,000 (roughly 5% of the total 1869 population (de la Fuente, 1872)). The indigenous population was mostly confined to these territories, which means that in practice the 1869 census largely excluded this population.

The US sample was constructed by following white males across the 1850 and 1880 full count US censuses, which are available through the North Atlantic Population Project (Ruggles et al., 2011). The sample is restricted to whites because slaves, who constituted the majority of the US black population at the time, were not individually listed in the 1850 population census. The 1850 census was restricted to the 31 states and 4 organized territories that comprised the US at the time.

To construct the British sample, I linked males across the 1851 and 1881 British censuses. These censuses included individuals residing in England, Wales, and Scotland. One challenge when constructing this sample is that the 1851 full count data (including names) is currently not publicly available. Hence, unlike in the other three countries, I did not link individuals from a full count to a full count census. Rather, I started from a 2% sample of the 1851 census and then linked individuals forward to the 1881 full count census. Starting from a sample might result in a higher rate of false positives than when linking a full count to a full count census.[5] The 1851 and 1881 British censuses

[5] The following example can illustrate why this might be the case: Assume there are two John Smiths in 1851 Britain, but only one of them is in the 2% sample. In 1860, one of the John Smiths (the one who was originally in the 2% sample) decides to move outside of Britain. By 1881, there will be just one John Smith in the census, so a linking method that starts from a sample will likely link the unique (in the 2% sample) 1851 John Smith to the unique 1881 full count John Smith, even though the two are different people.

are available from the UK Data Archive, as studies number 1316 and 4177-8, respectively.[6]

Finally, I constructed the Norwegian sample by linking males through the 1865 and 1900 full count Norwegian censuses. These two censuses are also available through the North Atlantic Population Project (Ruggles et al., 2011).

Given the absence of numerical individual-level identifiers, I linked individuals using information on reported names, place of birth and year of birth. This identifying information is non-unique and prone to enumeration and transcription errors. If the prevalence of errors or the selection into the sample differed across countries, it would be hard to disentangle true differences in mobility from differences in sample construction, a possibility I discuss in the robustness section. To alleviate this concern, I used a consistent linking methodology to create the four linked samples.

Specifically, I applied a procedure that is analogous to the one introduced by Winkler (1988) and first used in economic history by Mill and Stein (2012). The procedure has the following steps. In the first step, I identified a group of individuals in each of the earlier censuses that I would attempt to match to the later census. I then searched the later census for a set of potential matches for each of these individuals. I identified potential matches as individuals who: (1) reported the same place of birth, (2) had a predicted age difference of no more than five years in absolute value, and (3) had a first and last name starting with the same letter.[7] Based on the similarity of their reported names and predicted years of birth, I used the EM algorithm (Dempster et al., 1977), a standard tool in statistics, to calculate a linking score ranging from 0 to 1 for each pair of potential matches: higher scores represented pairs of records that were more similar to each other. To measure similarity in first and last names, I used the Jaro-Winkler string distance function (Winkler, 1990), whereas to measure similarity in ages I used the absolute value of the predicted years of birth.[8]

I used these linking scores to inform the decision rule on which records to incorporate to the analysis. To be considered a unique match for an individual in the earlier census, a record in the later census had to satisfy three conditions: (1) being the record with the highest linking score p_1

[6] Long and Ferrie (2013) only included English and Welsh-born males when creating the British 1851-1881 linked sample, since the 1881 Scottish data was not available at the time their analysis was conducted. The results (not reported) are similar if I restrict the British sample to English and Welsh-born males.

[7] The process of only comparing individuals who agree on certain characteristics is called "blocking" and is intended to reduce the number of comparisons that need to be made.

[8] Place of birth corresponds to provinces in Argentina, states in the US, parishes in Britain and municipalities in Norway. For the foreign born in each of these countries, there is no consistent information on birthplace beyond country of birth.

among all the potential matches for that individual, (2) having a linking score above a threshold ($p_1 > \underline{p}$, with $\underline{p} \in (0, 1)$), and (3) having a linking score sufficiently higher than the second highest linking score ($p_2 < l$, with $l \in [0, \underline{p})$). Further details on this method are provided in Abramitzky et al. (2018).

In addition to the algorithm that was used to link individuals across census years, these samples share two other important similarities. First, each of the samples was constructed by linking national censuses of population. Hence, all the samples include individuals from a diverse set of areas within each of the countries, as well as those who migrated internally. Second, the samples are limited to father-son pairs in which the son coresided with his father at the time of the initial census. This restriction biases all the samples towards intact households in the initial census year.

After linking the data for each of the countries, I imposed two additional sample restrictions. First, I only included sons who were 16 years old or less when observed living with their father in the initial census year. This adjustment corrects for the fact that individuals who coreside with their father until relatively late might exhibit different patterns of mobility than those who do not (Xie and Killewald, 2013). Second, I restricted the samples to father-son pairs in which both the father and the son were between the ages of 30 and 60 when their occupations were recorded in the census. This adjustment deals with the fact that occupations measured either too early or too late in the life cycle might be a noisy measure of long-run economic status (Solon, 1992). I emphasize, however, that none of these restrictions affects the conclusions of the paper. The final baseline samples include about 12,000 father-son pairs for Argentina, 180,000 for the US, 2,500 for Britain and 18,000 for Norway.

Because the linking is based on names, one concern is that the samples might not be representative of the underlying populations of interest. To investigate this issue, I first compare the occupational structure of fathers in the linked samples to that in the cross sections. To do so, I focus on males aged 30 to 60 years old who have a son who is 16 years old or less in the initial census year. For Argentina, I compare the linked sample to the 2.5% sample of the 1869 census constructed by Somoza (1967). For the US and Norway, I compare the linked samples to the cross-sectional samples available through the North Atlantic Population Project. For Britain, I compare the linked samples to the 2% sample of the 1851 census discussed above.

Table A.1 in the Online Appendix shows the distribution of fathers across occupational groups in the baseline year, in both the cross sections and the linked samples. The occupational structure in the linked samples is relatively similar to that in the cross sections for each of the countries, although some differences exist. In all four countries, the linked data overrepresent farmers relative to the remaining occupational groups. In Argentina, the US and Norway, this overrepresentation comes mainly at the expense of unskilled workers, whereas in Britain it comes mainly at the expense of skilled/semi-skilled workers.[9]

In tables A.2 to A.5 in the Online Appendix, I explore the extent to which individuals in the linked sample differed from those in the cross section with respect to characteristics other than the occupational category of the father. To do so, I estimate a Probit model of the probability of being in the matched sample. I focus on a set of variables that are consistently available across the four countries: father's and son's age, occupational category of the father, and region of residence in the initial census year. The tables report the marginal effects of each of these characteristics on the linkage probability. Younger individuals were more likely to be linked in Argentina, while the opposite holds for Norway. In the US and Argentina, individuals with older fathers are more likely to be linked, but no such correlation is apparent for Britain and Norway. In Argentina, the data oversample individuals who lived outside of Buenos Aires in the initial census year, likely reflecting that the information on names is less unique for individuals born in larger provinces. A similar pattern holds for those initially in London in the British data, as well as for those in the Eastern region of Norway (which includes Oslo). Finally, the US data oversample individuals from the Northeast relative to the South and the Midwest. Below, I show that the results are similar when reweighting the samples to account for selection with respect to these observable characteristics.

Classifying Occupations

Self-reported occupation is the only economic outcome that is consistently available across all of the countries and census years.[10] I classified occupations into four broad categories using the HISCLASS (Leeuwen et al., 2002) classification scheme: white-collar, farmer, skilled-semi skilled and unskilled.

[9] It is worth noting that the linked sample might differ from the baseline year cross-section for reasons unrelated to the linking procedure. For instance, selective mortality and outmigration might also result in the over/underrepresentation of some occupational groups.

[10] Other potentially relevant outcomes are included in just a subset of the datasets, which prevents me from using them in a systematic comparison. For instance, there is question on wealth in the US 1850 census, but no such question exists in the other datasets. Similarly,there is a question on literacy in both the Argentine and US samples, but no such questions are available in the 1851 British and 1865 Norwegian census.

To do so, I first assigned each occupation a code from the Historical International Classification of Occupations (HISCO). I then mapped each HISCO code to an occupational category using HISCLASS.[11] In the robustness section of the paper, I show that the results are similar when using less coarse occupational categories.

There are two important aspects to note about the farming category. First, the censuses do not include consistent information on ownership that would enable me to distinguish tenant farmers from owners. For instance, the question on ownership was included in the 1895 census of Argentina but was not included in 1869. Similarly, the 1850 US census contains a question on the value of real estate property that could be used to infer ownership status, but the 1880 census lacks such a question. This limitation of the data implies that it is not possible to produce a systematic mobility comparison that takes into account intergenerational changes in farm ownership status. Second, farm laborers are excluded from the farming category and are categorized as unskilled workers. This exclusion implies that the farmer category just encompasses those who independently operated a farm, either as owners or as tenants.

Measuring Occupational Mobility

The analysis of occupational mobility can be based on transition matrices in which rows represent fathers' occupations and columns represent sons' occupations. A simple summary measure of mobility based on these matrices is the fraction of sons who end up in a different occupational category than their father.

As discussed in Altham (1970) and Long and Ferrie (2013), a shortcoming of using this measure to compare mobility across countries is that it does not distinguish whether differences in mobility are due to: differences in the distribution of occupations across matrices, or differences in the strength of the association of the rows and columns of the matrices. An example can illustrate this point. Assume that we would like to compare mobility in countries P and Q, and that there are just two occupational categories. In country P, 50% of individuals work in occupational category 1 and 50% work in occupational category 2, both in the father's and the son's generation. In country Q, 25% of individuals work in occupational category 1 and 75% work in occupational category 2,

[11] I collapsed the HISCLASS scheme into four broad categories white-collar (HISCLASS 1-5), farmer (HISCLASS 8), skilled/semi-skilled (HISCLASS 6-7,9) and unskilled (HISCLASS 10-12)

both in the father's and the son's generation. Assume that in both countries the occupation of a son is *independent* of the occupation of his father. In this case, the fraction of sons working in a different occupational category than their father will on average be $\frac{1}{2}$ in country P, but only $\frac{3}{8}$ in country Q. Note, however, that in this hypothetical example there is no association between fathers' and sons' outcomes in either of the countries.[12]

With just two occupations, a more fundamental measure of mobility can be constructed based on the relative odds of accessing a given occupation for individuals with fathers in different occupations. Under perfect mobility, the relative odds are one: a father's occupation does not provide any advantage in accessing a given occupation. In this context, a natural measure of mobility is given by how far the odds ratio is from being one. In the previous example, the odds of children of category 1 workers of landing job 1 instead of job 2, relative to the odds of children of category 2 workers are: $\frac{\frac{1/2}{1/2}}{\frac{1/2}{1/2}} = 1$ in country P, and $\frac{\frac{1/4}{3/4}}{\frac{1/4}{3/4}} = 1$ in country Q. That is, both countries exhibit perfect mobility in this example.

With more than two occupational categories, each matrix has more than one possible cross-product ratio. For example, if there were three occupational categories instead of two, there would be nine cross-product ratios: One such product ratio is $\frac{p_{11}/p_{13}}{p_{21}/p_{23}}$, which represents the ratio between: (1) the odds that children of fathers in category 1 work in category 1 rather in category 3, and (2) the odds that children of fathers in category 2 work in category 1 rather in category 3.[13]

Hence, comparing the full mobility patterns in the case with more than two occupational categories requires aggregating the differences in cross-product ratios into a single statistic. To this purpose, I followed Long and Ferrie (2013) and used the Altham (1970) statistic. Given two matrices P and Q of size $r \times s$, the Altham statistic $d(P,Q)$ is given by:

$$d(P,Q) = \Big[\sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{m=1}^{r} \sum_{l=1}^{s} \Big[\log\Big(\frac{p_{ij}p_{lm}}{p_{im}p_{lj}}\Big) - \log\Big(\frac{q_{ij}q_{lm}}{q_{im}q_{lj}}\Big) \Big]^2 \Big]^{1/2} \tag{1}$$

where i and l index fathers' occupations and j and m index sons' occupations. This statistic

[12] $Mobility^p = \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} = \frac{1}{2}$ and $Mobility^q = \frac{1}{4} \times \frac{3}{4} + \frac{3}{4} \times \frac{1}{4} = \frac{3}{8}$

[13] More generally, when comparing two matrices each with n occupational categories, there are $\frac{n^2(n-1)^2}{4}$ independent cross product ratios. The intuition for the formula is the following: Each cross product ratio involves a pair of categories for the father and a pair categories for the son. With n categories, there are $\binom{n}{2}$ possible pairs of father's occupations. For each of these combinations, there are $\binom{n}{2}$ possible pairs of occupation for each children. Hence, there are $\binom{n}{2} \times \binom{n}{2} = \frac{n^2(n-1)^2}{4}$ possible cross product ratios.

aggregates the (log) differences between the cross-product ratios in matrix P (given by $\frac{p_{ij}p_{lm}}{p_{im}p_{lj}}$) and the cross-product ratios in matrix Q (given by $\frac{q_{ij}q_{lm}}{q_{im}q_{lj}}$).

Higher values of $d(P,Q)$ imply larger differences in the row-column association of matrices P and Q, but are not informative regarding which of the matrices exhibits more mobility. The next step is then to calculate the statistic $d(P,J)$ for each of the countries. This statistic measures the same difference in row-column association but relative to a matrix J representing independence (a matrix in which all the odds ratios are one). Higher values of $d(P,J)$ imply greater departures from independence (that is, less mobility). Hence, if we observe that $d(P,Q) \neq 0$ and that $d(P,J) < d(Q,J)$, we can conclude that P shows greater mobility than Q.

There are three useful properties of the Altham statistic. First, it is possible to perform a likelihood-ratio test to assess whether the statistic is significantly different from zero. Hence, it is possible to establish if the mobility levels of two countries are statistically different from each other. Second, the statistic can be decomposed into the different elements of the sum (based on the formula in equation 1). Hence, it is possible to assess which precise odds ratios explain most of the difference between the two matrices. Third, the measure treats movements across categories symmetrically regardless of the origin and destination categories. In other words, there is no assumed "distance" nor order between the occupational groups. This property is useful in this context since, given the absence of individual-level earnings data, it would be challenging to rank the occupational groups unambiguously.

Main Results

Table 1 shows the matrices of father-son occupational transitions in 1869-1895 Argentina, 1850-1880 US, 1851-1881 Britain and 1865-1900 Norway. Rows in these matrices represent the occupation of a father in the initial year, and columns represent the occupation of his son in the end year. The last row of each matrix shows the occupational distribution among sons in the end year.

The table shows some well-known facts about the occupational structure of these countries in the second half of the nineteenth century. The most salient difference across countries is the much lower proportion of fathers employed in farming in Britain. By 1851, the first stage of the British industrial revolution had been mostly completed. Indeed, Crafts (1984) argues that the most striking feature of the British industrial revolution was the "very rapid and early change in the share of the labor force in the primary sector". In contrast, the proportion of fathers employed as farmers was similarly high in Norway (51%) and Argentina (49%), and even higher in the US (65%). Although the relative importance of farming had in all cases declined, it was still the predominant occupation in the son's generation in these three countries (last row in each of the matrices).

I next turn to the analysis of intergenerational movements across occupational categories. To do so, in table 2, I provide summary measures of mobility for each of the countries. The first such measure is the fraction of sons in a different occupational category than their father, reported in panel (a) of table 2. About 55% of the sons in the Argentine sample worked in a different occupational category than their father, compared to 45% in the US, 44% in Britain, and 45% in Norway. This simple metric suggests higher levels of mobility in Argentina than in every other country included in the comparison.

However, as discussed above, this metric has the limitation of not being able to distinguish between differences in the degree of intergenerational association and differences in the occupational structure across countries. Hence, in panel (b) of table 2, I present the results of calculating the Altham statistics for each of the countries in the data. Each element of the table represents the Altham statistic comparing country i to country j.[14] In the first column, I present the Altham statistic comparing each country to a transition matrix representing full independence.

First, for each pair of countries I reject the hypotheses that their mobility patterns are the same. The smallest value of the statistic is the one corresponding to the Argentina-US comparison (4.23), suggesting that mobility patterns were the closest among these two countries. Indeed, as can be seen from the top two panels of table 1, the mobility matrices of Argentina and the US are quite similar to each other. For instance, in both countries a similar fraction of the children of white-collar workers became white-collar workers themselves (53% in both countries) and a similar fraction of these children became farmers instead (22% in Argentina, 21% in the US). This pattern suggests that the similarity in mobility levels of the two countries is not driven by the use of the Altham statistic as the measure of mobility.

Next, to rank countries in terms of mobility, I measure the distance of each of them with

[14]Note that the Altham statistic can be interpreted as a distance metric and hence satisfies $d(i, j) = d(j, i)$.

respect to a matrix representing independence. First, note that for all of the countries I reject the hypotheses that their mobility patterns are the same that would arise under independence (all the Altham statistics in column 1 are significantly different from zero). In terms of the ranking of countries, the data suggest that mobility in Argentina was slightly higher than in the US; the distance with respect to independence in Argentina is 13.45 compared to 14.67 in the US. Mobility in Argentina and the US was much higher than in Britain and Norway (20.89 and 25.94, respectively). The departure from independence is about twice as large in Norway and about 50% larger in Britain relative to Argentina and the US. The results also suggest that mobility was higher in Britain than in Norway. Overall, this pattern is consistent with mobility levels being higher in the "settler economies" of the New World than in Europe.[15]

Robustness of Main Results

I first analyze the robustness of the results to using less coarse occupational categories. Table A.6 in the Online Appendix shows that the results are qualitatively similar when I use five instead of four occupational categories. In the first panel of this table, I re-estimate the Altham statistic splitting the white-collar category into "high white-collar" (professional, technical, and kindred; managers, officials, and proprietors) and "low white-collar" (clerical and sales). In the second panel, I split the unskilled category into those employed as farm laborers and the remaining unskilled workers. As expected, the departure from independence is higher in all of the matrices, and especially in the Norwegian case. However, the ranking of countries in terms of distance with respect to independence does not change in either of the panels (column 1). Similarly, all the differences in mobility across countries remain large and statistically significant.[16]

An additional concern is that the results might be driven by the procedure used to create the linked samples. In table 3, I show the robustness of the main results to features of the linking procedure. First, I redo the mobility comparisons using a more conservative linking procedure *only*

[15] My estimates of the Altham statistic for the US, Britain and Norway are similar to those reported in previous papers. Long and Ferrie (2013) find an Altham statistic of 22.7 for Britain and of 11.9 for the US. Long and Ferrie (2018) more recent study on three-generational mobility finds a father-son Altham statistic of 23.9 for Britain and of 14.49 for the US. Modalsli (2017) finds an Altham statistic of 24.1 for 1865-1900 Norway.

[16] There is likely some misclassification when dividing unskilled workers into those employed as farm laborers and the rest. The reason is that some workers simply reported "laborers" as their occupation, without specifying the sector of the economy in which they were employed. Some of these workers were likely farm laborers, but the census data do not enable me to identify them with precision.

for the Argentine and US samples. In this way, I bias the sample towards finding less mobility in Argentina and the US. To do so, I first restrict the Argentine and US samples to exact matches with respect to age and name. Second, because individuals with common names are more likely to be incorrectly linked (Bailey et al., 2017), I exclude observations in the top 25% in terms of first name frequency within their province/state of birth, and then observations in the top 50% of first name frequency. Rows 2 to 4 of table 3 show that these exercises result in a very similar pattern to the one that I obtain in the baseline analysis (row 1).

Yet, it is theoretically possible that the fraction of false positives in the Argentine and US samples was higher than in the British and Norwegian samples even when using a relatively more conservative sample just for the former two countries. How much larger would the rate of false positives in Argentina and the US need to be to generate the observed mobility differences with respect to Britain and Norway?

The observed mobility matrices for each of the countries are a combination of: the "true" transition matrices (when father-son links are correct) and the "false" transition matrices (when father-son links are incorrect). Assume that the transition matrix among unrelated father-sons is equal to the transition matrix that we would observe under independence (which would be the case if the linking algorithm randomly matched sons to fathers).

Denote the fraction of false positives in the linked sample corresponding to country A as α_A. Under the above assumption, the observed distance with respect to independence –$d(A, J)$– can be written as:

$$d((1 - \alpha_A)A^* + \alpha_A J, J) \tag{2}$$

where A^* is the true (unobserved) transition matrix in country A.

To fix ideas, assume that we would like to compare the mobility levels of Argentina (A) and Britain (B). Further, assume that: (1) the rate of false positives in the British data were zero ($\alpha_B = 0$), and that (2) the true mobility levels were the same in both countries (which will be true if $A^* = B^*$). This exercise hence gives a *lower bound* on the rate of false positives that we would need in the Argentine data to explain the observed difference in mobility with respect to Britain. This lower bound is given by $\underline{\alpha}_A$ such that:

$$d((1-\underline{\alpha}_A)B+\underline{\alpha}_A J,J)=d(A,J) \quad (3)$$

where I made use of the assumption that $A^* = B^* = B$ (that is, that there are no errors in the British data and that Argentina and Britain have the same true levels of mobility). Also, note that if we replace the right hand side of equation 3 with $d(U, J)$, we can use this expression to find the value $\underline{\alpha}_U$ that would rationalize the observed differences in mobility between Britain (B) and the US (U), under the assumption of no true differences in mobility between the two countries.

Panel (a) of figure 1 shows the value of $d((1-\alpha)B+\alpha J,J)$ (y-axis) for different values of α (x-axis), where B corresponds to the observed mobility matrix of Britain. The intersection between $d((1-\alpha)B+\alpha J,J)$ and the observed value of $d(A,J)$ hence corresponds to $\underline{\alpha}_A$, the lower bound in the share of false positives in the Argentine data that we would need to rationalize the observed difference in mobility with respect to Britain. Similarly, the intersection between $d((1-\alpha)B+\alpha J,J)$ and $d(U,J)$ corresponds to $\underline{\alpha}_U$, the lower bound in the share of false positives in the US data. Panel (b) shows the corresponding figure for the Argentina-Norway and US-Norway comparisons.

Panel (a) shows that we would need the rate of false positives in the Argentine data to be at least 40% to rationalize the observed differences between Argentina and Britain. Panel (b) shows that, if Argentina and Norway had the same true levels of mobility, we would need the rate of false positives in the Argentine sample to be at least 44% to explain the observed differences. Similarly, for the US to have the same mobility as Britain, we would need the rate of false positives in the US data to be at least 33%, and for the US and Norway to have the same levels we would need this rate to be of at least 38%. While it is not possible to theoretically discard such high levels of false positives in the Argentina and US samples relative to Britain and Norway, it seems unlikely that samples that were generated using similar sources and the same linking procedure would exhibit such large differences in the rate of false positives.[17]

[17] It is also possible to compute the lower bound of the rate of false positives in the Argentine data that would rationalize the observed differences with respect to the US, under the assumption of no true differences in mobility. I find that a rate of false positives that is just 4 percentage points higher for Argentina than for the US can rationalize the observed differences in mobility. Similarly, it is possible to find the lower bound of the rate of false positives in the British data that would rationalize the observed British-Norway differences. I find that a 15 percentage points higher rate of false positives in the British data can rationalize the observed differences with respect to Norway. These figures are reported in Online Appendix figure A.1. Note that in both cases, the lower bounds are considerably smaller than the lower bounds corresponding to the Argentina-US versus Britain-Norway comparison, which is why I emphasize the latter results.

The second concern related to the linking procedure is that the results might be driven by differential selection into the linked samples. First, note that the Altham statistic already corrects for differences in the size of occupational categories across countries. So, even if the samples were biased towards individuals with fathers in certain occupational categories, the value of the Altham statistic would be correct provided that the selection was not on *mobility*. Second, note that this differential selection would only be a challenge to my main findings if the linked samples oversampled individuals with higher *mobility* prospects in Argentina and the US. While there are no strong a priori reasons to expect this pattern, I cannot fully rule out this possibility since I only observe mobility for individuals in the linked sample. Yet, to partially address this concern, in the last row of table 3 I report estimates of the Altham statistic after reweighting the data to account for selection on observables into the linked sample. To do so, I use the inverse of the linkage probabilities estimated in tables A.2 to A.5 as weights. The results indicate a very similar pattern to the one that I obtain with the baseline (unweighted) samples.

Why was Mobility Higher in the US and Argentina than in Britain and Norway?

The results above indicate that mobility levels were higher in Argentina and the US than in Britain and Norway. In addition, Long and Ferrie (2013) show that mobility differences between the US and Britain had largely disappeared by the second half of the 20th century. Hence, any explanation of this set of results is likely related to features of the economies of Argentina and the US that distinguished them from Britain and Norway in the late 19th century, but for which such differences did not exist in the second half of the 20th century.

Decomposing Differences in Mobility

As a first step towards understanding the economic sources of differences in mobility, I decompose the Altham statistics corresponding to the Argentina-US, Argentina-Britain, and Argentina-Norway comparisons into their different components based on the formula in equation 1. This decomposition enables me to quantify the relative importance of each odds ratio in explaining the mobility differences across countries. The further apart the odds ratio in one country is from the odds ratio in the other, the larger its contribution to the Altham statistic comparing both countries.[18] Hence, in searching for economic explanations for the observed differences between countries, a natural starting point will be to focus on factors related to components of the Altham statistic that explain a high fraction of these differences.

Table 4 shows the results of these decompositions. In each case, I report the five (out of 36) largest elements of the statistic. In the top panel, I present the comparison of Argentina and Britain, while in the bottom two panels I present the Argentina-Norway and Argentina-US comparisons, respectively. For each of the components, I also report the corresponding odds ratio in each of the countries.

The main difference between Argentina and Britain corresponds to the odds at which sons of farmers entered farming rather than unskilled work relative to the sons of unskilled workers. In Argentina, the sons of farmers were 2.5 more likely to enter farming relative to unskilled work than the sons of unskilled workers. This same magnitude takes a much larger value (above 25) for Britain. The second largest component corresponds to the odds at which the children of farmers entered farming rather than skilled/semi-skilled work relative to the children of skilled/semi-skilled workers. Panel (a) of table A.7 in the Online Appendix shows that the difference between the US and Britain stemmed from similar sources as the difference between Argentina and Britain. Indeed, the largest two elements are the same in both comparisons. Overall, these results are consistent with the main difference between Argentina and Britain being that entering farming was relatively easier for the children of non-farmers in Argentina. I note, however, that because the Altham statistic is based on *odds ratios*, it is not possible to establish whether these differences were driven by easier access to farming for children of non-farmers or by less attachment to farming among children of farmers (or a combination of both).

The main difference between Argentina and Norway stems from the odds at which children of white-collar workers entered white-collar jobs rather than farming compared to children of farmers. In Argentina, this ratio took a value of approximately 9, whereas in Norway this ratio was above 120. The second largest difference corresponds to the odds at which children of white collar workers entered white collar jobs rather than farming compared to children of unskilled workers. This ratio

[18] Because each odds ratio is based on a pair of father's occupations and a pair of son's occupation, there is no straightforward way of weighting the odds ratio based on the size of the occupational categories that it involves.

took a value of about 5 in Argentina and of more than 40 in Norway. This pattern suggests that entering white-collar jobs was relatively easier for the children of non-white collar workers in Argentina than in Norway (or that entering farming was relatively harder). In contrast to the Argentina-Britain comparison, not all of the major components of the Altham statistict comparing Argentina and Norway involve the farming category. For instance, the relative odds at which sons of white collar workers entered white-collar rather than unskilled occupations relative to the children of unskilled workers was 11 in Argentina, but more than 60 in Norway. Panel (b) of table A.7 in the Online Appendix also shows a similar pattern when comparing the US and Norway.

Where did the small differences between Argentina and the US stem from? The results in the bottom panel of table 4 suggests that the relative chances of children of farmers and children of unskilled workers were closer in Argentina than in the US (first three rows of panel (c)). However, note that the odds ratios are in all cases not too different from each other when comparing these two countries.

Mobility Differences in Historical Context

The analysis above suggests that differences in the likelihood of entering farming for children of non-farmers was an important source for the higher mobility in Argentina and the US. One likely reason for this relative ease was the abundance of land in Argentina and the US compared to Britain and Norway.

Argentina and the US had abundant productive land and a sparse population by the mid 19th-century. By 1869, Argentina had a population of about 2 million and a population density of just 2 persons per square mile. Similarly, the US had a population of 23 million and a population density below 8 persons per square mile. By contrast, population density was above 13 persons per square mile in Norway and close to 300 persons per square mile in Britain.

The expansion of the frontier further increased the relative land abundance of Argentina and the US during this period. From 1869 to 1895, Argentina incorporated about 600,000 square miles of land to the control of the central government, out of which an important fraction was located in the fertile plains of the country (Conde, 1979). Similarly, from 1850 to 1880 the US continued a process of westward expansion: even until the late 19th century, the US contained large tracts of open land (Ferrie, 1997; Bazzi et al., 2017).

Land abundance per se would not translate into opportunities for "newcomers to the land" if this land was not accessible for independent small-scale farming. In this regard, there were important differences between how public lands were allocated in Argentina and in the US. In Argentina, land grants typically allocated large plots of lands to relatively few individuals. By contrast, land grants greatly facilitated the ownership of small plots of land in the US. In particular, the Homestead Act of 1862 made small plots essentially free for families willing to settle and work the land for a specified period (Engerman and Sokoloff, 2008).

Although allocation of public lands in Argentina did not favor the access to small plots as much as in the US, a relatively recent literature in history shows an important economic role for small scale independent farming in this period (Hora, 2007).[19] This literature emphasizes how the expansion of the agricultural frontier provided "ample horizon of opportunity for newcomers to the land", and resulted in a dynamic rural sector (Hora, 2001). Indeed, about half of the individuals that I classified as farmers in the 1895 census owned real estate property (a strong indication that they might have owned the land they worked). Moreover, as discussed in Hora (2001), access to independent farming did not require ownership in a frontier economy with high land abundance. In the Argentine context, tenants exploited plots of similar size, made similar investment decisions and had similar productivity levels as land owners (Adelman et al., 1994; Taylor, 1997).

The available historical evidence (based on land register data) also suggests that the picture of a relatively fluid rural sector began to change by the turn of the century in Argentina. Some authors have attributed this decline in mobility to the closing of the agricultural frontier. For instance, Hora (2001) writes that: "At the same time, frontier expansion came to an end and the rural social structure became increasingly rigid". Losada (2009) notes that the Argentine economy at the turn of the century featured: "the gradual closing of opportunities for upward mobility offered by a frontier economy in its initial stages of development".

It is important to note, however, that the high mobility of Argentina and the US in this time period was not limited to families who initially resided in "frontier" areas of both countries. In figure 2, I compute the Altham statistic separately for each Argentine province and US state, based on the location of a family in the initial census year. Lighter colors represent lower values of the

[19] When describing the current state of the historiography of Argentine agriculture in the 19th century, Hora (2007) writes (my own translation from Spanish): "Currently, small and medium agricultural producers (mostly family enterprises), are perceived as decisive factors in the process of agricultural surplus generation."

Altham statistic (that is, higher levels of mobility). In Argentina, mobility levels were the highest in the Center and West regions of the country, and the lowest in the North. In the US, mobility levels were the highest in the Midwest and the lowest in the South, while the Northeast exhibited intermediate levels of mobility. But even in the least mobile US state, the distance with respect to independence was lower than in Britain and Norway. Similarly, mobility levels in Argentina were higher than in Britain and Norway in all but the two northernmost provinces of the country. In other words, while there were regional heterogeneities, high levels of mobility were a relatively widespread phenomenon in Argentina and the US.

This relative ease in accessing land for independent farming in Argentina and the US contrasts with the situation in Britain and Norway. In Britain, the vast majority of agricultural land was already privately owned by 1850 (Clark and Clark, 2001). Moreover, access to independent farming was unlikely to be an important avenue of mobility as the size of the primary sector was already quite small by the mid 19th-century. In Norway, the fraction of the workforce in farming was similar to Argentina and the US. However, land scarcity resulted in the growth of the cottager population, and is cited as one of the main reasons for the mass emigration of Norwegians to the US (Milward and Saul, 2013).

Beyond land abundance, another important similarity between Argentina and the US in the second half of the 19th century (and arguably, a likely consequence of this abundance) is that both countries attracted large numbers of European immigrants. Argentina and the US were the two largest destination countries in the Age of Mass Migration, receiving about 6 and 30 million immigrants, respectively (Hatton and Williamson, 1998). Immigrant families might exhibit higher rates of intergenerational mobility if first-generation immigrants suffer a penalty in the labor market that is then eroded by the second generation.[20] Hence, another plausible explanation for the higher mobility in Argentina and the US is the higher fraction of immigrants in these two countries.

To test whether these compositional effects may explain the higher mobility in Argentina and the US, I recompute the Altham statistic excluding the father-son pairs in which the father is an immigrant from the Argentine and US samples. If the difference between Argentina and the US and Europe was just driven by differences in the fraction of immigrants in the population, we should

[20] For instance, Feigenbaum (2017) finds higher intergenerational mobility among the grandchildren of immigrants in early 20th-century Iowa.

observe similar mobility levels across countries when excluding foreigners from the analysis. The second row in table 5 shows that a higher fraction of immigrants cannot account for the higher mobility in Argentina and the US. First, excluding immigrants only leads to a small change in the distance of the mobility matrices of Argentina and the US with respect to independence (first and second rows of columns 1 and 2). Second, the difference with respect to both Britain and Noway remains of similar size and statistically significant for both Argentina and the US (first and second rows of columns 6 to 9).

This analysis does not imply that mass migration had no effect on mobility, The inflow of large numbers of foreigners may have shook the social structures of the receiving countries, and might have been an additional reason for the higher levels of mobility in Argentina and the US. However, this analysis shows that migration-related compositional effects cannot account for the higher mobility levels in these two countries.

The other side of mass migration to the Americas is that both Britain and Norway had a much larger share of *emigrants* than Argentina and the US. One possible explanation for the differences in mobility would be a sample selection issue: the most mobile Old World sons moved to the New World, resulting in overall less mobility among stayers, but perhaps similar mobility overall. It is possible to bound the quantitative importance of this factor. To do so, I assume that the occupation of sons who migrated internationally was independent of the occupation of their fathers –that is, that emigrating sons experienced the maximum possible amount of mobility.

Denote the transition matrix among British stayers as B (an analogous reasoning applies for the Norwegian matrix). Let J be a matrix representing the predicted transitions under independence. For simplicity, I assume that the occupational distribution of fathers among stayers is equal to the occupational distribution of fathers among emigrants. Under these assumptions, the simulated transition matrix including emigrants and stayers is given by:

$$(1 - \alpha)B + \alpha J \tag{4}$$

where α is the fraction of emigrants in the population. This matrix is a weighted average between the observed transition matrix and the predicted transition matrix under independence.

The third row in table 5 shows the results of comparing the observed mobility matrices of

Argentina and the US to the simulated matrices of Britain and Norway, under the assumption of perfect mobility among European emigrants. In the baseline exercise, I assume that the share of sons who emigrated out of Britain was 10%, and that the share of sons who emigrated out of Norway was 20%.[21]

Overall, this exercise suggests that emigration out of Europe cannot fully account for the observed differences in mobility. Even under the extreme assumption of perfect mobility among European emigrants, the difference with respect to Argentina and the US remains large and statistically significant. Indeed, for mobility in Britain and Norway to be similar to that in Argentina and the US, the emigration rate would have to be above 40% (which is considerably higher than the observed rates of outmigration), and all of these individuals be perfectly occupationally mobile.

Finally, I briefly discuss three explanations for the observed differences in mobility that are ex-ante plausible but do not seem entirely consistent with the data. The first of these explanations is differences in the progressivity of public investments in human capital, as emphasized in Solon (2004). In the 19th century, the US was at the forefront of a global rise in mass public schooling (Lindert, 2004): 68.1 percent of 5-14 year olds attended primary school in 1850 US, compared to 49.8 percent in 1851 Britain and 46.6 in 1865 Norway. However, only 16% of individuals in this age range attended school in 1869 Argentina. This pattern is at odds with the observed differences in mobility being driven by higher public investments in human capital in Argentina and the US. In any case, it is not even clear that these early investments in public schooling disproportionately favored the poor in these countries (Parman, 2011).

A chief explanation for current cross-country differences in intergenerational mobility is that mobility levels are lower in countries where inequality is higher (Krueger, 2012). I am not aware of any research comparing these four countries with respect to late 19th-century inequality. However, the scattered available evidence does not appear entirely consistent with a simple inequality story. First, Lindert and Williamson (2016) document similar levels of inequality in the US and Britain in this time period. Second, Williamson (2015) documents similar levels of income inequality in Latin America –although the analysis does not include Argentina – and in the rest of the Western world in the period. I note, however, that a full comparison of inequality across these four countries is

[21] Table 1 in Taylor and Williamson (1997) reports the cumulative labor force emigration rate for a number of countries in the Atlantic economy. For Britain, the cumulative rate from 1870 to 1910 (a 40 years period) was 11%. For Norway, the cumulative rate from 1870 to 1910 was 24%.

not possible for this time period given the absence of individual-level earnings data.

The last explanation relates to differences in fertility rates across countries. One important change that was taking place in the mid-19th century was the fertility transition. If fertility rates were lower in Europe, then each family could allocate its resources among fewer offspring, perhaps facilitating the transmission of economic status across generations. Indeed, Britain was at the forefront of the fertility transition in this time period (Guinnane, 2011). However, even when restricting the samples of Argentina and the US to households with below median fertility rates, the results still indicate higher mobility in Argentina and the US than in Britain and Norway (last row of table 5).[22]

Conclusions

Prominent theories of comparative economic development (Sokoloff and Engerman, 2000) point to persistent inequality and lack of social mobility as a source of the relatively poor performance of Latin American economies. Yet, there is little empirical evidence on how historical levels of social mobility in Latin America compared to those of other regions of the World. In this paper, I compared rates of intergenerational occupational mobility across four countries in the late 19th-century: Argentina, the US, Britain and Norway. To the best of my knowledge, this paper constitutes the first international comparison of 19th-century social mobility that uses consistent data and methodology and includes a Latin American or currently developing country.

The results indicate that Argentina and the US had higher intergenerational occupational mobility than Britain and Norway in the late 19th century. One factor that likely played a role in explaining these differences is the higher availability of unexploited land in Argentina and the US relative to Britain and Norway. This feature of the economies of Argentina and the US also characterized other "settler economies" such as Australia and Canada, suggesting that mobility rates might have also been high in these economies in this time period. In other words, there was likely nothing "exceptional" about the high levels of mobility of 19th-century US.

[22]Because families in which the father was younger will mechanically have fewer children at the time of the census due to incomplete fertility, I focus on the sample of households in which the father was at least 40 years old in the baseline year. The results (available upon request) are also similar if I exclude first-born sons from the sample, thus suggesting that differences in inheritance practices are also an unlikely explanation for the observed differences in mobility.

The high levels of social mobility in late 19th-century Argentina posit a historical and theoretical puzzle. Benabou and Ok (2001) and Piketty (1995) argue that high levels of mobility would tend to dampen demands for income redistribution. Indeed, the ethos of the *American Dream* has been used to rationalize why the US engages in less redistribution than other developed countries (Alesina and La Ferrara, 2005; Alesina et al., 2018). So, why did Argentina, despite featuring high levels of mobility in the 19th-century, end up engaging in much higher levels of redistribution than the US over the course of the 20th-century? Further research is needed to understand this issue.

References

Abramitzky, R., R. Mill, and S. Pérez (2018): "Linking Individuals Across Historical Sources: a Fully Automated Approach," Tech. rep., National Bureau of Economic Research.

Adelman, J. et al. (1994): *Frontier development: land, labour, and capital on the wheatlands of Argentina and Canada, 1890-1914*, Oxford University Press.

Alesina, A. and E. La Ferrara (2005): "Preferences for redistribution in the land of opportunities," *Journal of public Economics*, 89, 897–931.

Alesina, A., S. Stantcheva, and E. Teso (2018): "Intergenerational mobility and preferences for redistribution," *American Economic Review*, 108, 521–54.

Altham, P. M. (1970): "The Measurement of Association of Rows and Columns for an r× s Contingency Table," *Journal of the Royal Statistical Society. Series B (Methodological)*, 63–73.

Bailey, M., C. Cole, M. Henderson, and C. Massey (2017): "How Well Do Automated Linking Methods Perform in Historical Samples? Evidence from New Ground Truth," Tech. rep., Working Paper.

Bazzi, S., M. Fiszbein, and M. Gebresilasse (2017): "Frontier Culture: The Roots and Persistence of" Rugged Individualism" in the United States," Tech. rep., National Bureau of Economic Research.

Benabou, R. and E. A. Ok (2001): "Social Mobility and the demand for redistribution: The POUM Hypothesis." *Quarterly Journal of Economics*, 116.

Chen, Y., S. Naidu, T. Yu, and N. Yuchtman (2015): "Intergenerational mobility and institutional change in 20th century China," *Explorations in Economic History*, 58, 44–73.

Chetty, R., N. Hendren, P. Kline, and E. Saez (2014a): "Where is the land of opportunity? The geography of intergenerational mobility in the United States," *The Quarterly Journal of Economics*, 129, 1553–1623.

Chetty, R., N. Hendren, P. Kline, E. Saez, and N. Turner (2014b): "Is the United States still a land of opportunity? Recent trends in intergenerational mobility," *The American Economic Review*, 104, 141–147.

Cilliers, J., J. Fourie, et al. (2016): "Social mobility during South Africaâ™ s industrial take-off," *ERSA (Economic Research Southern Africa) Working Paper*, 617.

Clark, G. (2014): *The son also rises: surnames and the history of social mobility*, Princeton University Press.

Clark, G. and A. Clark (2001): "Common rights to land in England, 1475–1839," *The Journal of Economic History*, 61, 1009–1036.

Conde, R. C. (1979): "El progreso argentino: 1880-1914," .

Crafts, N. F. (1984): "Patterns of development in nineteenth century Europe," *Oxford Economic Papers*, 36, 438–458.

de la Fuente, D. G. (1872): *Primer censo de la República Argentina: Verificado en los dias 15, 16 y 17 de setiembre de 1869*, Impr. del Porvenir.

De Tocqueville, A. (1835): *Democracy in america*, vol. 10, Regnery Publishing.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977): "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the royal statistical society. Series B (methodological)*, 1–38.

Engerman, S. L. and K. L. Sokoloff (2008): "Once upon a time in the Americas: land and immigration policies in the New World," in *Understanding Long-Run Economic Growth: Geography, Institutions, and the Knowledge Economy*, University of Chicago Press, 13–48.

Feigenbaum, J. J. (2017): "Multiple Measures of Historical Intergenerational Mobility: Iowa 1915 to 1940," *mimeo.*

Ferrie, J. (1997): "Migration to the Frontier in Mid-Nineteenth Century America: A Re-Examination of Turner's "Safety Valve"," .

Guinnane, T. W. (2011): "The historical fertility transition: A guide for economists," *Journal of Economic Literature*, 49, 589–614.

Hatton, T. J. and J. G. Williamson (1998): *The age of mass migration: Causes and economic impact*, Oxford University Press on Demand.

Hora, R. (2001): *The landowners of the Argentine pampas: A social and political history 1860-1945*, Clarendon Press.

——— (2007): "La evolución de la desigualdad en la Argentina del siglo XIX: una agenda en construcción," *Desarrollo Económico*, 487–501.

Jonsson, J. O., D. B. Grusky, R. Pollak, M. Di Carlo, and C. Mood (2011): "Occupations and social mobility: gradational, big-class, and micro-class reproduction in comparative perspective," .

Krueger, A. (2012): "The rise and consequences of inequality," *Presentation made to the Center for American Progress, January 12th. Available at http://www. americanprogress. org/events/2012/01/12/17181/the-rise-and-consequences-of-inequality.*

Leeuwen, M. v., I. Maas, and A. Miles (2002): *HISCO: Historical international standard classification of occupations*, Leuven: Leuven University Press.

Lindert, P. H. (2004): *Growing public: Volume 1, the story: Social spending and economic growth since the eighteenth century*, vol. 1, Cambridge University Press.

Lindert, P. H. and J. G. Williamson (2016): *Unequal Gains: American Growth and Inequality since 1700*, Princeton University Press.

Long, J. and J. Ferrie (2007): "The path to convergence: intergenerational occupational mobility in Britain and the US in three eras," *The Economic Journal*, 117.

——— (2013): "Intergenerational occupational mobility in Great Britain and the United States since 1850," *The American Economic Review*, 103, 1109–1137.

——— (2018): "Grandfathers Matter (ed): Occupational Mobility Across Three Generations in the US and Britain, 1850-1911," *The Economic Journal*.

Losada, L. (2009): *Historia de las elites en la Argentina: desde la conquista hasta el surgimiento del peronismo*, SciELO Argentina.

Marx, K. (1852): *The 18th Brumaire of Louis Bonaparte*, Wildside Press LLC.

Mill, R. and L. C. Stein (2012): "Race, Skin Color, and Economic Outcomes in Early Twentieth-Century America," Tech. rep., Working Paper, Stanford University December.

Milward, A. and S. B. Saul (2013): *The Economic Development of Continental Europe 1780-1870*, Routledge.

Modalsli, J. (2017): "Intergenerational Mobility in Norway, 1865–2011," *The Scandinavian Journal of Economics*, 119, 34–71.

Olivetti, C. and M. D. Paserman (2015): "In the name of the son (and the daughter): Intergenerational mobility in the United States, 1850–1940," *The American Economic Review*, 105, 2695–2724.

Parman, J. (2011): "American mobility and the expansion of public education," *The Journal of Economic History*, 71, 105–132.

Piketty, T. (1995): "Social mobility and redistributive politics," *The Quarterly journal of economics*, 110, 551–584.

Ruggles, S., E. Roberts, S. Sarkar, and M. Sobek (2011): "The North Atlantic population project: Progress and prospects," *Historical methods*, 44, 1–6.

Sokoloff, K. L. and S. L. Engerman (2000): "History lessons: Institutions, factors endowments, and paths of development in the new world," *The Journal of Economic Perspectives*, 14, 217–232.

Solon, G. (1992): "Intergenerational income mobility in the United States," *The American Economic Review*, 393–408.

——— (2002): "Cross-country differences in intergenerational earnings mobility," *Journal of Economic Perspectives*, 16, 59–66.

——— (2004): "A model of intergenerational mobility variation over time and place," *Generational income mobility in North America and Europe*, 38–47.

——— (2015): "What do we know so far about multigenerational mobility?" Tech. rep., National Bureau of Economic Research.

Somoza, J. L. (1967): *Muestras de los dos primeros censos nacionales de población, 1869 y 1895*, 46, Instituto Torcuato Di Tella, Centro de Investigaciones Sociales.

Taylor, A. M. (1997): "Latifundia as Malefactor in Economic Development? Scale, Tenancy, and Agriculture on the Pampas, 1880-1914," .

Taylor, A. M. and J. G. Williamson (1997): "Convergence in the age of mass migration," *European review of economic history*, 1, 27–63.

Williamson, J. G. (2015): "Latin American inequality: colonial origins, commodity booms, or a missed 20th century leveling?" Tech. rep., National Bureau of Economic Research.

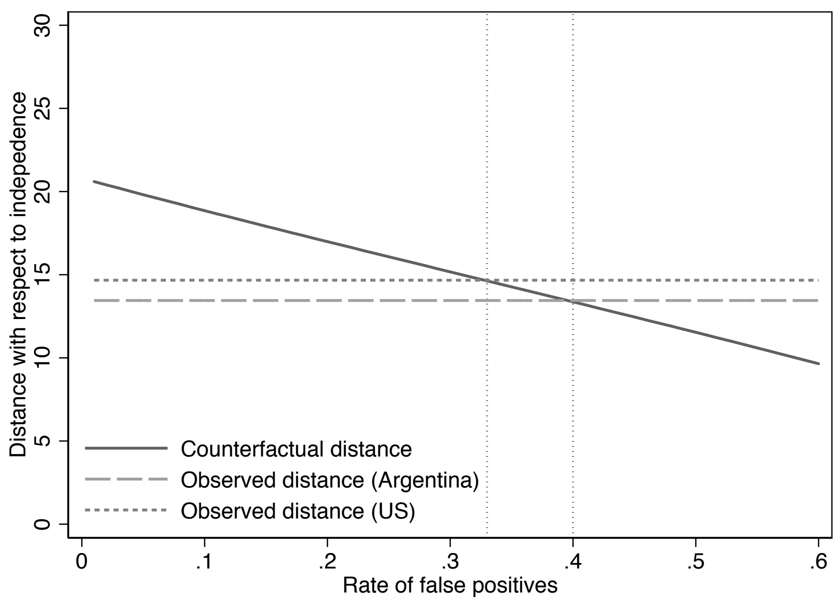
Winkler, W. E. (1988): "Using the EM algorithm for weight computation in the Fellegi-Sunter model of record linkage," in *Proceedings of the Section on Survey Research Methods, American Statistical Association*, vol. 667, 671.

——— (1990): "String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage." .

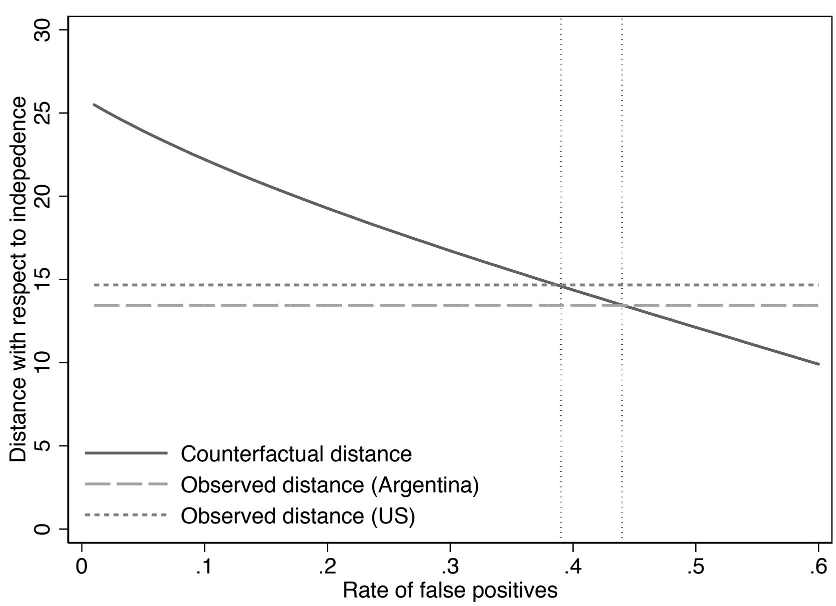
Xie, Y. and A. Killewald (2013): "Intergenerational occupational mobility in Great Britain and the United States since 1850: Comment," *The American economic review*, 103, 2003–2020.

Figure 1: Robustness to Higher Share of False Positives in the Argentine and US Data

(a) Britain



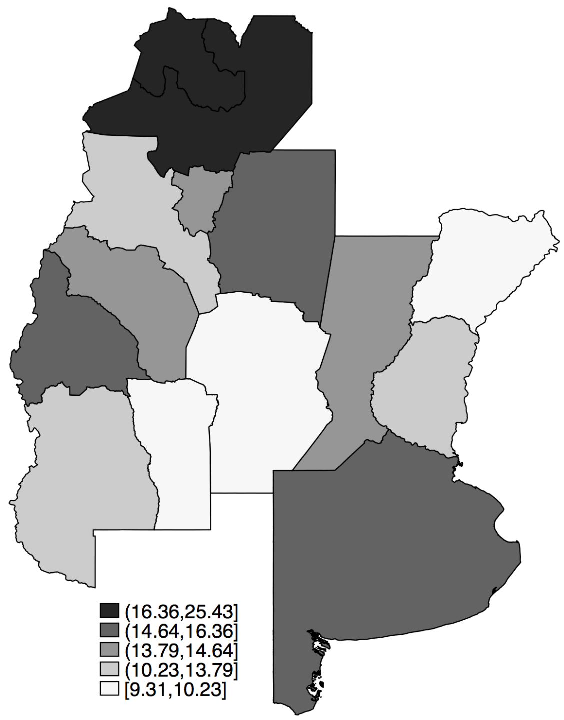
(b) Norway



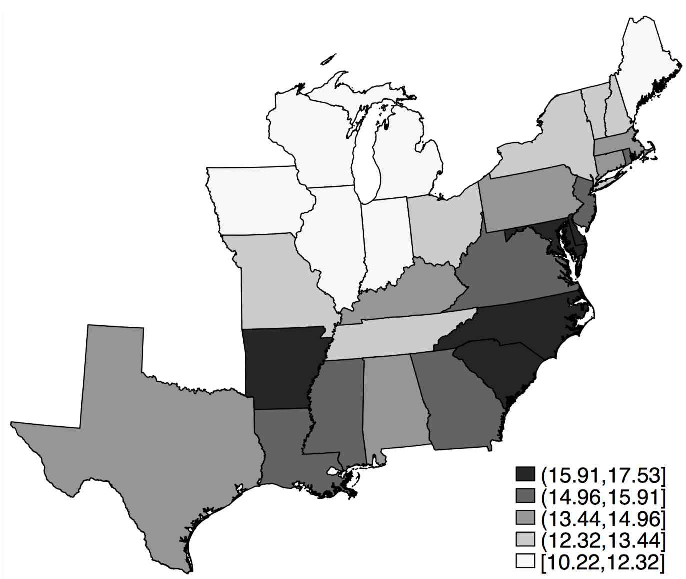
Notes: This figure shows the rate of false positives that would be needed to eliminate the observed differences in mobility between Argentina and the US and Britain and Norway. The counterfactual distance with respect to independence is constructed under the assumptions that: (1) the rate of false positives is zero for both Britain and Norway, and (2) there are no true differences in mobility between Argentina and the US and Britain and Norway. See main text for further details.

Figure 2: Altham statistic by Initial Province/State of Residence in Argentina and the US

(a) Argentina



(b) United States



Notes: This figure shows the Altham statistic (distance with respect to a matrix representing full independence) computed separately for each province/state, based on a family place of residence in the initial census year. Lighter colors represent lower values of the Altham statistic (higher mobility).

Table 1: Intergenerational Occupational Mobility. Row Percent (Frequencies)

	Son's occupation				
Father's occupation	White collar	Farmer	Skilled/semi-skilled	Unskilled	Row total
Argentina 1869-1895					
White-collar	**0.53**	0.22	0.12	0.13	1
	(**995**)	(411)	(234)	(245)	(1885)
Farmer	0.13	**0.51**	0.10	0.26	1
	(803)	(**3083**)	(586)	(1598)	(6070)
Skilled/semi-skilled	0.29	0.24	**0.28**	0.20	1
	(564)	(470)	(**539**)	(387)	(1960)
Unskilled	0.15	0.31	0.14	**0.40**	1
	(363)	(727)	(335)	(**958**)	(2383)
Column total	0.22	0.38	0.14	0.26	1
	(2725)	(4691)	(1694)	(3188)	(12298)
US 1850-1880					
White-collar	**0.53**	0.21	0.19	0.08	1
	(**7798**)	(3094)	(2792)	(1129)	(14813)
Farmer	0.13	**0.62**	0.13	0.12	1
	(15282)	(**72296**)	(15470)	(13785)	(116833)
Skilled/semi-skilled	0.23	0.24	**0.41**	0.12	1
	(7881)	(8374)	(**14084**)	(4312)	(34651)
Unskilled	0.13	0.29	0.32	**0.25**	1
	(1601)	(3512)	(3903)	(**3067**)	(12083)
Column total	0.18	0.49	0.20	0.12	1
	(32562)	(87276)	(36249)	(22293)	(178380)
Britain 1851-1881					
White-collar	**0.46**	0.05	0.39	0.10	1
	(**151**)	(18)	(127)	(33)	(329)
Farmer	0.19	**0.37**	0.25	0.19	1
	(53)	(**103**)	(68)	(53)	(277)
Skilled/semi-skilled	0.17	0.03	**0.71**	0.10	1
	(180)	(27)	(**770**)	(103)	(1080)
Unskilled	0.12	0.03	0.39	**0.45**	1
	(102)	(29)	(335)	(**384**)	(850)
Column total	0.19	0.07	0.51	0.23	1
	(486)	(177)	(1300)	(573)	(2536)
Norway 1865-1900					
White-collar	**0.80**	0.05	0.11	0.05	1
	(**1455**)	(84)	(191)	(84)	(1814)
Farmer	0.09	**0.62**	0.14	0.15	1
	(813)	(**5799**)	(1325)	(1454)	(9391)
Skilled/semi-skilled	0.30	0.06	**0.52**	0.13	1
	(640)	(129)	(**1116**)	(277)	(2162)
Unskilled	0.10	0.24	0.30	**0.36**	1
	(481)	(1211)	(1473)	(**1801**)	(4966)
Column total	0.18	0.39	0.22	0.20	1
	(3389)	(7223)	(4105)	(3616)	(18333)

Notes: This table shows the intergenerational occupational transition matrices for Argentina (1869-1895), the US (1850-1880), Britain (1851-1881) and Norway (1865-1900). Rows represent the occupation of the father in the initial census year. Columns represent the occupation of the son in the final census year.
Sources: Father-son linked samples for Argentina, the US, Britain and Norway as described in the main text.

Table 2: Summary Measures of Intergenerational Occupational Mobility

(a) Fraction in Different Occupational Category than Father

Argentina	US	Britain	Norway
0.55	0.45	0.44	0.45

(b) Altham Statistics

	Independence	Argentina	US	Britain	Norway
Independence	.	.	.	.	.
Argentina	13.45***	.	.	.	.
US	14.67***	4.23***	.	.	.
Britain	20.80***	11.09***	9.35***	.	.
Norway	25.94***	13.61***	12.73***	14.00***	.

Notes: Panel (a) reports the fraction of sons working in a different occupational category than their father, based on the occupational transition matrices in table 1. Each element of panel (b) shows the Altham statistic corresponding to a pair of transition matrices: a matrix representing independence, Argentina, US, Britain and Norway. Larger values of the statistic correspond to larger differences in the row-column association of the matrices. Larger differences with respect to independence represent lower mobility. For each of these distances, I performed a test of the hypothesis that $d(i, j) = 0$. Significance levels are indicated by $* * *p < 0.01$, $* * p < 0.05$, $*p < 0.1$.

Table 3: Robustness: Higher Intergenerational Occupational Mobility in Argentina and the US than in Britain and Norway

	d(Arg.,Ind.) (1)	d(US,Ind.) (2)	d(Arg.,Brit.) (3)	d(Arg.,Nor.) (4)	d(US,Brit.) (5)	d(US,Nor.) (6)
Baseline	13.45	14.67	11.09	13.61	9.35	12.73
Exact links	15.38	15.73	10.59	12.29	9.00	12.35
Excl. top 25% names	13.35	15.16	12.28	13.85	9.31	12.38
Excl. top 50% names	13.35	14.75	11.67	13.70	9.43	12.69
Reweigthed sample	13.53	14.65	11.16	13.61	9.56	12.77

Notes: All the Altham statistic are significantly different from zero at the 1% level. This table shows the robustness of the finding of higher intergenerational occupational mobility in 19th-century Argentina and US than in Britain and Norway. In the "Exact links" row, I restrict the Argentine and US samples to records that match exactly in terms of their identifying information. In the "Excl. top 25% frequent names" row, I exclude from the Argentine and US samples those individuals with a first name in the top 25% in terms of commonness within their province/state of birth. In the "Excl. top 50% frequent names" row, I exclude from the Argentine and US samples those individuals with a first name in the top 50% in terms of commonness within their province/state of birth. in the "Reweighted sample" row, I reweight the data to account for selection in observables into the linked sample using the weights estimated in Online Appendix tables A.2 to A.5.

Table 4: Components of Differences in Intergenerational Occupational Mobility

(a) Argentina versus Britain

Contrast	Odds ratio		Percent of total
	Argentina	Britain	
$[FF/FU]/[UF/UU]$	2.542	25.733	0.174
$[FF/FS]/[UF/US]$	2.424	17.497	0.127
$[FF/FS]/[SF/SS]$	6.033	43.197	0.126
$[FF/FU]/[SF/SU]$	1.589	7.414	0.077
$[WF/WS]/[FF/FS]$	0.334	0.094	0.053
Top 5			0.558

(b) Argentina versus Norway

Contrast	Odds ratio		Percent of total
	Argentina	Norway	
$[WW/WF]/[FW/FF]$	9.295	123.551	0.145
$[WW/WF]/[UW/UF]$	4.849	43.610	0.104
$[FW/FF]/[SW/SF]$	0.217	0.028	0.090
$[FF/FS]/[SF/SS]$	6.033	37.863	0.073
$[WW/WU]/[UW/UU]$	10.718	64.856	0.070
Top 5			0.481

(c) Argentina versus US

Contrast	Odds ratio		Percent of total
	Argentina	US	
$[FF/FS]/[UF/US]$	2.424	5.194	0.130
$[FW/FS]/[UW/US]$	1.265	2.408	0.093
$[FF/FU]/[UF/UU]$	2.542	4.580	0.077
$[FW/FU]/[SW/SU]$	0.345	0.607	0.071
$[WW/WS]/[UW/US]$	3.924	6.809	0.068
Top 5			0.439

Notes: The first letter denotes the occupation of the father and the second letter denotes the occupation of the son. W=White-collar, F=Farmer, S=Skilled/semi-skilled, U=Unskilled. In panel (a), I show the contribution of each component of the Altham statistic comparing Argentina and Britain. In panels (b) and (c), I repeat the analysis for the Altham statistic comparing Argentina and Norway, and Argentina and the US, respectively. In each of the cases, I report the five (out of 36) largest components of the Altham statistic, sorted in decreasing order.

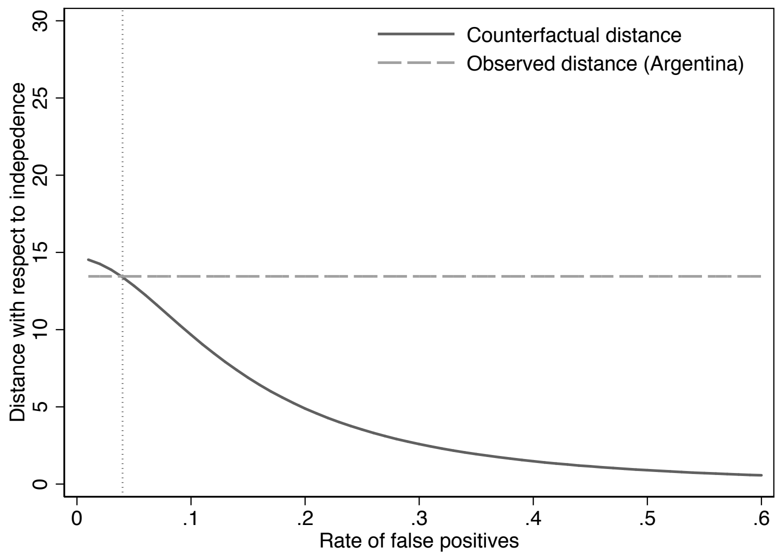
Table 5: Explaining Differences in Intergenerational Occupational Mobility

	d(Arg.,Ind.) (1)	d(US,Ind.) (2)	d(Brit.,Ind.) (3)	d(Nor.,Ind.) (4)	d(Arg.,US) (5)	d(Arg.,Brit.) (6)	d(Arg.,Nor.) (7)	d(US,Brit.) (8)	d(US,Nor.) (9)
Baseline	13.45	14.67	20.80	25.94	4.23	11.09	13.61	9.35	12.73
Immigrants	12.24	14.75	20.80	25.94	4.17	11.90	15.22	9.57	12.54
Emigrants	13.45	14.67	18.85	19.28	4.23	9.61	7.08	7.84	6.30
Low fertility	13.50	14.50	20.80	25.94	4.61	10.90	13.45	9.53	13.17

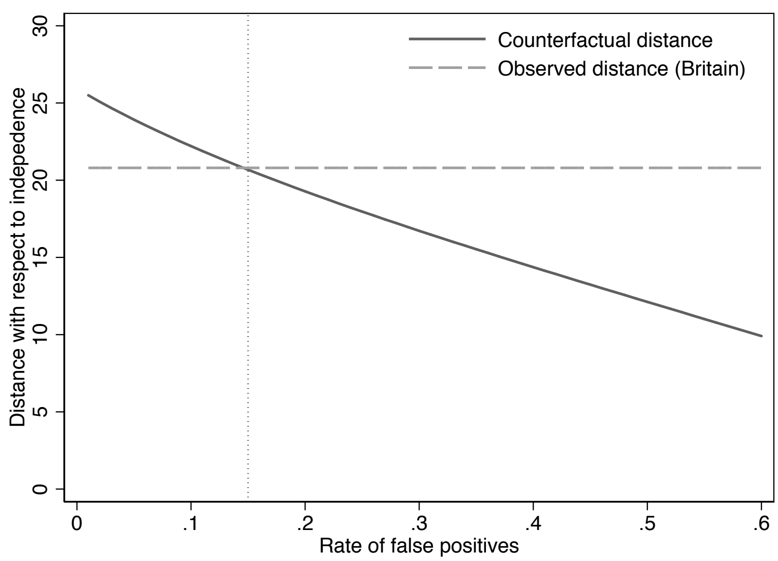
Notes: All the Altham statistic are significantly different from zero at the 1% level. In the "Immigrants" row, I recompute the Altham statistics after excluding children of immigrants from the Argentina and US samples. In the "Emigrants" row, I recompute the Altham statistics after replacing the Britain and Norway observed matrices with a counterfactual matrix assuming perfect mobility among emigrants from these two countries. See the main text for details and assumptions on this exercise. In the "low fertility" row, I restrict the Argentina and US samples to households with below median fertility.

Figure A.1: Counterfactual Distance with Respect to Independence

(a) Argentina vs US differences



(b) Britain vs Norway differences



Notes: This figure shows the rate of false positives that would be needed to eliminate the observed differences in mobility between Argentina and the US (panel (a)) and between Britain and Norway (panel (b))., The counterfactual distance with respect to independence is constructed under the assumptions that: the rate of false positives is zero for the country with the lowest measured mobility in each pair and there are no true differences in mobility between each pair of countries. See main text for further details.

Table A.1: Comparing the Linked Samples to the Population. Occupational Structure of Fathers

	White-collar	Farmer	Skilled/semi-skilled	Unskilled
i. Argentina, 1869				
Cross-section	0.12	0.47	0.20	0.22
Panel	0.16	0.49	0.16	0.19
ii. US, 1850				
Cross-section	0.08	0.59	0.22	0.10
Panel	0.08	0.65	0.19	0.07
iii. Britain, 1851				
Cross-section	0.14	0.08	0.48	0.30
Panel	0.13	0.11	0.43	0.33
iv. Norway, 1865				
Cross-section	0.06	0.48	0.13	0.33
Panel	0.10	0.51	0.12	0.27

Notes: This table shows a comparison between the occupational structure of fathers in the cross sectional data and in the linked sample. For Argentina, I use the linked sample as described in the main text and the census sample constructed by Somoza (1967). For the US, and Norway, I use the census samples available through the North Atlantic Population Project (Ruggles et al., 2011). For Britain, I use the 2% sample of the 1851 census as described in the main text.

Table A.2: Probit Marginal Effects on Linkage, Argentina 1869

	(1) 1 if in linked sample
Demographic	
Age	-0.009***
	(0.0009)
Father's age	0.003***
	(0.0004)
Father's Occupation	
White-collar	0.097***
	(0.0107)
Farmer	0.029***
	(0.0082)
Skilled/semi-skilled	-0.025**
	(0.0099)
Residence	
East	0.071***
	(0.0084)
North	0.053***
	(0.0079)
West	0.109***
	(0.0111)
Observations	25213

Notes: This table shows the marginal effects of a probit model estimating the probability that an observation belongs to the linked sample when linking the 1869-1895 Argentine censuses. The cross-sectional data corresponds to the 1869 census and is from Somoza (1967). The omitted category are children of unskilled workers from the province of Buenos Aires.

Table A.3: Probit Marginal Effects on Linkage, US 1850

	(1) 1 if in linked sample
Demographic	
Age	-0.000
	(0.0002)
Father's age	0.001***
	(0.0001)
Father's Occupation	
White-collar	0.057***
	(0.0037)
Farmer	0.070***
	(0.0028)
Skilled/semi-skilled	0.039***
	(0.0031)
Residence	
Midwest	-0.008***
	(0.0019)
South	-0.022***
	(0.0019)
Observations	214570

Notes: This table shows the marginal effects of a probit model estimating the probability that an observation belongs to the linked sample when linking the 1850-1880 US censuses. The cross-sectional data corresponds to the 1850 census and is from IPUMS (Ruggles et al., 2011). The omitted category are children of unskilled workers from the Northeast.

Table A.4: Probit Marginal Effects on Linkage, Britain 1851

	(1) 1 if in linked sample
Demographic	
Age	-0.000
	(0.0002)
Father's age	0.000
	(0.0001)
Father's Occupation	
White-collar	-0.002
	(0.0031)
Farmer	0.013***
	(0.0035)
Skilled/semi-skilled	-0.007***
	(0.0022)
Residence	
London	-0.028***
	(0.0047)
Midlands-East	0.002
	(0.0030)
North	-0.010***
	(0.0032)
South	0.016***
	(0.0032)
Wales	-0.020***
	(0.0049)
Observations	51090

Notes: This table shows the marginal effects of a probit model estimating the probability that an observation belongs to the linked sample when linking the 1851-1881 British censuses. The cross-sectional data corresponds to the 1851 census and corresponds to the UK Data Archives study number 1316. The omitted category are children of unskilled workers from Scotland.

Table A.5: Probit Marginal Effects on Linkage, Norway 1865

	(1) 1 if in linked sample
Demographic	
Age	0.001***
	(0.0001)
Father's age	0.000
	(0.0001)
Father's Occupation	
White-collar	0.052***
	(0.0021)
Farmer	0.014***
	(0.0012)
Skilled/semi-skilled	0.011***
	(0.0018)
Residence	
North	0.038***
	(0.0017)
Trondelag	-0.004*
	(0.0019)
West	0.003***
	(0.0013)
South	0.010***
	(0.0019)
Observations	256064

Notes: This table shows the marginal effects of a probit model estimating the probability that an observation belongs to the linked sample when linking the 1865-1900 Norwegian censuses. The cross-sectional data corresponds to the 1865 census and is from IPUMS (Ruggles et al., 2011). The omitted category are children of unskilled workers from the Eastern region.

Table A.6: Altham Statistics with Five Occupational Categories

(a) Low and high white-collar

	Independence	Argentina	US	Britain	Norway
Independence	.	.	.	.	.
Argentina	23.34***	.	.	.	.
US	27.97***	14.85***	.	.	.
Britain	32.64***	19.11***	19.82***	.	.
Norway	44.70***	26.65***	21.87***	26.07***	.

(b) Unskilled and farm laborers

	Independence	Argentina	US	Britain	Norway
Independence	.	.	.	.	.
Argentina	22.27***	.	.	.	.
US	30.45***	15.70***	.	.	.
Britain	32.43***	19.81***	23.65***	.	.
Norway	37.11***	20.59***	18.66***	25.50***	.

Notes: This table presents a version of panel (b) of table 2 using five occupational categories instead of four. In panel (a), I split the white-collar category into "low white-collar" and "high white-collar". In panel (b), I split the unskilled category into "farm laborers" and the remaining unskilled workers.

Table A.7: Components of Differences in Intergenerational Occupational Mobility

a. US versus Britain

Contrast	Odds ratio		Percent of total
	US	Britain	%
$[FF/FU]/[UF/UU]$	4.580	25.733	0.136
$[FF/FS]/[SF/SS]$	7.860	43.197	0.133
$[FF/FS]/[UF/US]$	5.194	17.497	0.067
$[SS/SU]/[US/UU]$	2.567	8.569	0.066
$[FW/FF]/[UW/UF]$	0.464	0.146	0.061
Top 5			0.464

b. US versus Norway

Contrast	Odds ratio		Percent of total
	US	Norway	%
$[WW/WF]/[FW/FF]$	11.923	123.551	0.135
$[FW/FF]/[SW/SF]$	0.225	0.028	0.106
$[WW/WF]/[UW/UF]$	5.529	43.610	0.105
$[SW/SF]/[UW/UF]$	2.064	12.491	0.080
$[WW/WU]/[FW/FU]$	6.230	30.978	0.063
Top 5			0.490

Notes: The first letter denotes the occupation of the father and the second letter denotes the occupation of the son. W=White-collar, F=Farmer, S=Skilled/semi-skilled, U=Unskilled. In the top panel , I show the contribution of each element of the Altham statistic comparing the US and Britain. In the bottom panel, I repeat the analysis for the Altham statistic comparing the US and Norway. In each of the cases, I report the five (out of 36) largest elements of the Altham statistic, sorted in decreasing order.