Working Paper No. 329

Bias in Social Mobility Estimates with Historical Data:
Evidence from Swiss Microdata

Giacomin Favre

July 2019

Bias in Social Mobility Estimates with Historical Data
Evidence from Swiss Microdata

Giacomin Favre[*]

July 18, 2019

Abstract

This paper explores a variety of potential issues one has to address when estimating intergenerational mobility with historical data. Many studies are potentially affected by bias originating from individuals emigrating and thus dropping out of the sample, missing information on the life-cycle, and imperfectly linking data sets. Unique panel data on Zurich's citizenry between 1799 and 1926 entail information on true intergenerational links, and allow to follow individuals across the globe and time. This information enables me to explore how father-son mobility estimates are affected by excluding emigrating individuals, occupational patterns over the life-cycle, and linking procedures. The results suggest that focusing on geographically immobile individuals might decrease the estimated level of social mobility. The estimated level of mobility depends on both the father's and the son's age at classification but does not exhibit a monotone trend in the direction of the bias. Most recent linking procedures do not generate significant bias in the sample of Zurich citizens due to the high level of detail of the data combined with a small population size.

Keywords: Social Mobility, Geographic Mobility, Life-cycle, Matching, Historical Data.
JEL-Classification: J62, J61, N33, N34

1 Introduction

Does everyone have an equal chance of being economically successful? Or is socioeconomic status transmitted from one generation to the other such that the under-privileged are forever excluded from money and power? These questions are generally discussed in the literature on intergenerational mobility. Unique panel data of Zurich's citizenry between 1799 and 1926 allow me to contribute to the existing literature in another dimension by addressing the question: how large are the deviations in mobility estimates originating from geographic mobility, life-cycle patterns, and data linkage?

Usually, studies on social mobility depend on linking census data or birth registers to retrieve intergenerational links and thus obtain information on two or more generations (e.g. Ferrie, 2005, Bourdieu et al., 2009, Long, 2013, Long and Ferrie, 2013, Dribe et al., 2015, Barone and Mocetti, 2016, Collins and Wanamaker, 2017, Feigenbaum, 2018, Modalsli, 2017, Pérez, 2017). A similar procedure is necessary to follow individuals over the course of their life or to track emigrants to their destination country (e.g. Abramitzky et al., 2012, Abramitzky and Boustan, 2017). Thus, most studies omit emigrants and are not able to address potential life patterns in occupations.

The data at hand allow to estimate intergenerational mobility taking all of these potential sources of bias into consideration (as in Favre et al., 2018). Even more importantly, it is possible to quantify the magnitude and direction of the mentioned distortions as the employed data set includes true intergenerational links, allows me to track individuals over the course of their lives, and even provides information after emigration. I employ different measures of intergenerational mobility to control for the variety of measures that are employed in the existing literature. Hence, the main research questions in the paper are: (1) How are social mobility estimates affected by the in- and exclusion of migrants, life-cycle patterns in occupational outcomes, the linking procedure, and the employed measure of mobility, and (2) what is the relative size of the resulting deviations? By answering these questions, the paper also contributes to the literature on geographic mobility, life-cycle bias, and linking procedures.

The bulk of the literature on the economics of (international) migration (Constant and Zimmermann, 2013) tackles the assimilation of immigrants in their destination countries

which is directly linked to the intergenerational persistence of socioeconomic status.[1] Only a limited number of papers have focused on the selection of international migrants in their country of origin, and even fewer in a historical context. Notable exceptions are Wegge (1999, 2002, 2010), who finds that emigrants in mid-nineteenth century Germany were intermediately selected with respect to their socioeconomic positions, and Abramitzky et al. (2012, 2013), who provide evidence on negative selection among migrants from Norway to the United States during the second half of the nineteenth century.[2] This paper's contribution to the scarce literature on selection of international migrants in a historical context is twofold. First, I will provide suggestive evidence on the selection of international migrants in nineteenth century Switzerland and the effect on estimates of social mobility. Second, I investigate whether the selection of migrants differed across destination countries grouped by continent.

What are potential issues when employing non-longitudinal data sets? Income, wealth, occupation, and to some extent even education change over an individual's course of life. Hence, a person's socioeconomic status at one specific point in time might not be representative of the "lifetime" socioeconomic status. This results in life-cycle bias.[3] In this paper, I will investigate whether individuals also exhibit a life pattern with respect to occupational categories in a historical context. Further, I will explore whether implied father-son mobility differs across different ages at classification of both the father's and the son's generation.

Record linkage is widely applied in both historical and contemporaneous research contexts (Ruggles et al., 2018 provide an overview). Consequently, many researchers aim at automating and improving linking procedures[4] or evaluating existing mechanisms (Bailey et al., 2017a; Eriksson, 2017; Massey, 2017). The latter strand of literature compares different record linking procedures with "ground truth" data featuring the highest achiev-

---

[1]See e.g. Borjas (1992, 1993, 1994, 1995), Hammarstedt and Palme (2006), and Ward (2017) on historical migration waves or Card (2005) and Bauer and Riphahn (2007) on more recent migration waves.

[2]See Abramitzky and Boustan (2017) on a review of migration flows to the United States.

[3]Solon (1999) surveys some research on life-cycle bias in intergenerational mobility estimates. Life-cycle bias is found to be large and of varying direction (Jenkins, 1987; Grawe, 2006; Nybom and Stuhler, 2016a). Thus, one strand of the literature has focused on correcting for this bias (Haider and Solon, 2006; Böhlmark and Lindquist, 2006; Nybom and Stuhler, 2016b; Gregg et al., 2017).

[4]See e.g. Scheuren and Winkler (1993), Ferrie (1996, 2004), Christen and Goiser (2007), Herzog et al. (2007), Goeken et al. (2011), Baskerville et al. (2014), Abowd (2017), Bailey et al. (2017b), Abramitzky et al. (2012, 2014, 2019).

able matching rates. Both Bailey et al. (2017a) and Eriksson (2017) investigate the effect of record linkage on estimates of historical social mobility. They establish their benchmark "ground truth" data employing additional information on ancestors, which is not contained in other studies' data, resulting in matched data sets with higher quality. This paper, on the other hand, is the first to evaluate linking precision of linking procedures and the related bias in intergenerational mobility estimates in a historical context based on *observable* intergenerational links.

I find that Swiss emigrants were on average intermediately selected in the nineteenth century, and that this selection differed strongly by destination continent. Individuals moving to European countries were more positively selected than those moving to the United States or Canada. This selection translates into different estimates of intergenerational mobility. Emigrants were, on average, more mobile than geographically immobile individuals. Further, individuals in the data exhibited occupational life patterns suggesting that they experienced upward intragenerational mobility. The older a citizen, the more likely he was to obtain a higher socioeconomic position. These life patterns affect the point estimates of intergenerational mobility to some extent. However, these biases do not exhibit a monotone trend nor do the patterns agree across measures of occupational mobility. Finally, state-of-the-art automated linking procedures perform neatly. They are able to match around 77 to 95 percent of all father-son pairs with rates of false matches in the single-digit per mille range. This is mostly caused by the small sample size and the high quality of the data. Still, linked samples do not exhibit structurally different estimates of intergenerational mobility in this sample of Zurich citizens in the nineteenth century.

The remainder of this paper is structured as follows. Section 2 describes the employed data. The main results are presented in Section 3. Section 4 concludes.

2 Data and Descriptives

Data Source   The data originate from the same twenty-five editions of the directory of citizens of the city of Zurich between 1799 and 1926 as described in Favre et al. (2018). This source contains the universe of Zurich's adult male citizenry. Every edition of the directory of citizens includes references to individuals' direct male relatives. As citizenship is inherited for men (jus sanguinis), the son of a citizen is a citizen as well, irrespective of his place of residence. Hence, the data feature observable intergenerational links of the male lineages. The information on individuals includes the first name, middle names, the last name, the place of family origin[5], the year of birth, the year of death, the place of residence, occupations, the number of houses owned, the military rank, and public offices (e.g. member of the municipality council) for several points in time. Information on women is scarcer. Daughters of citizens are indirectly referred to through their father with name and year of birth (unmarried) or with information on the husband (married). Wives are also indirectly referred to through their husbands by name, year of birth, and place of origin.

The key features of the data I exploit in this paper are threefold. First, the data contain information on the occupation and place of residence of citizens living abroad. The information on emigrants was partially acquired through foreign authorities and partially by mail-in forms. If no current information was available, the directory contains the latest available characteristics including their date and a note that the corresponding citizen was currently untraceable.[6] Hence, the data allow to track citizens across the globe and to observe their occupation. Second, the frequent release of a new citizens' directory (every two to eleven years) and its cross-section character allow to track citizens over time. Thus, I am able to observe the life pattern of occupations for all citizens. Third, the data contain observable intergenerational links through the cross-reference to all male relatives of a citizen. Consequently, there is no need for linking fathers to sons in an automated procedure.

---

[5]This information conveys the origin of the family before it was naturalized in Zurich dating back to even before the thirteenth century.

[6]The data lack information on emigrant citizens in only 0.63 percent of all entries (across all observation years).

Classifications of Occupations   I employ the same set of occupational classifications as described in Favre et al. (2018). For the main part of the analysis, I divide occupations into low, middle, and high socioeconomic positions (SEP). This division is based on the categorization introduced by Schüren (1989) for occupations in nineteenth century Germany. The most prevalent occupations in the low SEP category are locksmith, mechanic, and baker. Merchants dominate the middle SEP category, but also engineer and teacher are frequent occupations with middle SEP. Lastly, the high SEP category's most frequent occupations are priest, physician, and professor. The upside of this categorization is that one can interpret the classes in an ordinal manner. Most of the mobility measures employed do not rely on this feature but the ordinal distinction of occupations allows to shed light on e.g. the selection of migrants in greater detail. Farmers are excluded throughout the entire analysis as their share in Zurich's citizenry is negligible.

For better international comparability, I employ an alternative categorization similar to Long and Ferrie (2013), Modalsli (2017), and Pérez (2017). Occupations are first divided into manual and non-manual labor. The manual workers are subsequently split into an unskilled workers group (requires little to no training) and into a skilled workers group (requires some training or education). The non-manual workers are classified as white-collar workers. In a more detailed classification the white-collar workers can be split into lower and higher managers.[7] This classification is based on the Historical International Standard Classification of Occupations (HISCO) according to Van Leeuwen et al. (2002) that allows to map occupations into HISCLASS (Historical International Social Class Scheme, Van Leeuwen and Maas, 2011).[8] Throughout the paper, this classification is referred to as Long-Ferrie (three groups) or extended Long-Ferrie (four groups) categorization.

Samples and Descriptives   To analyze the three potential sources of bias, the data have to be split accordingly. On the one hand, one needs to generate a baseline sample that should capture the "true" level of mobility. On the other hand, one has to get a hold of each source of bias through separate sub-samples. Below, I describe each of the

---

[7]The most frequent occupations per category are: priest, physician, and engineer (higher white-collar group), merchant, shop clerk, and innkeeper (lower white-collar group), mechanic, baker, and blacksmith (skilled workers), and mercenary, upholsterer, and glazier (unskilled workers).

[8]The HISCLASS groups are distributed across occupational categories as follows: HISCLASS 1–2 (higher white-collar), HISCLASS 3–5 (lower white-collar), HISCLASS 6–7 (skilled workers), HISCLASS 9–12 (unskilled workers), and HISCLASS 8 (farmers—omitted).

samples separately.

The baseline sample in this paper consists of all father-son pairs that are available in Zurich's citizenry between 1799 and 1926. I do not divide the sample across time as the focus does not lie on changes in the level of mobility over time but on the size and direction of bias. Every individual is categorized according to his occupation around the age of forty with respect to both the SEP and Long-Ferrie classifications.[9] Some descriptives on the baseline sample are provided in Table 1. There are over 11,000 father-son pairs whereby one father may have several sons. Due to the structure of the data, fathers are on average older than sons at the observed occupation. There are some fathers in the early directories of citizens that are already older than forty. Similarly, some sons are younger than forty in the last available directories. The distribution across occupational categories is similar across the two generations.

The investigation of bias through geographic mobility requires a split of the data according to the emigration status of individuals. In this paper, I focus on international geographic mobility of sons. Thus, sons are divided into geographically immobile ("Stay"), return migrants ("Return"), and emigrants ("Emigrated"). Immobile sons may move away from the city of Zurich but remain in Switzerland. Return migrants spend some years away from Switzerland but return thereafter. Emigrants migrate to a different country and stay abroad. Both return migrants and emigrants may migrate repeatedly. To shed more light on the selection of migrants, I further split the geographically mobile group (emigrants and return migrants) by destination continent (Europe, North America, South America, Africa, Asia, and Australia).[10] Table 2 describes the three broad migrant groups and Table 3 contains further details on individuals by destination.[11] Roughly, seventy percent of all sons never lived abroad. The remaining thirty percent leave the country, and just over eleven percent remain abroad (emigrants).

One has to categorize individuals at different ages in order to evaluate how the age at classification affects estimates of intergenerational mobility. Thus, I categorize sons according to their occupation around twenty, thirty, and forty and fathers according to their occu-

---

[9]I exclude all individuals that are younger than sixteen or older than sixty-five.

[10]As only few individuals moved to South America, Africa, Asia, and Australia, I combine the four in one group (SA/Afr/As/Aus). Migrants are categorized into more than one destination continent if they migrate repeatedly.

[11]Table 30 in Appendix A.2 splits the group emigrating to South America, Africa, Asia, and Australia.

pation around thirty, forty, and fifty.[12] Different from the baseline sample, I only allow for a deviation from the specific classification age by five years.[13] This allows to construct mobility measures for father-son pairs for every combination of son's and father's age at the observed occupation. In order not to encounter issues with comparability, I exclude every father-son pair that lacks an observation around any of the corresponding ages. Table 4 entails descriptives of the resulting sample of 1,899 father-son pairs.

Lastly, the evaluation of automated linking mechanisms requires splitting the data into fathers and sons by ignoring the observable intergenerational link and re-matching the two. First, I construct a "fathers sample" that contains all of the information on the father and only the first name, last name, and year of birth of the son. Second, I construct a "sons sample" that contains all sons as a pool of potential matches with information on the first name, middle names, last name, year of birth, and the family's place of origin. Third, I employ automated linking procedures to join the fathers with their conjectural sons. There exist many possible mechanisms to perform this kind of record linkage (for a review, see e.g. Ruggles et al., 2018). To narrow the focus down to two of the most promising linking methods, I follow the recommendations by Bailey et al. (2017a) and evaluate the mechanism introduced by Ferrie (1996) and the one developed by Abramitzky et al. (2012, 2014).[14] In general, both mechanisms rely on a similar procedure. Based on first name, last name, (implied) age, and state of birth, they link individuals across time. In order to correct for orthographic differences in the spelling of names, both (may) employ phonetic corrections (NYSIIS, Soundex, or None)[15]. The basic steps of Ferrie (1996) can be condensed as follows: (1—optional) correct names phonetically, (2—optional) truncate first name after fourth letter, (3) match and discard if not born in same state, (4) allow for age differences of up to two years among matches, and (5) choose matched link with smallest difference in age. In this application, I employ the family's place of origin instead of state of birth and observe the year of birth rather than the age of individuals. The

---

[12]I choose this set of classification ages by generation to balance the remaining sample size and age spread.

[13]So, an individual that is categorized around the age of forty has to be between thirty-five and forty-five.

[14]Abramitzky et al. (2014) provide their code on `https://ranabr.people.stanford.edu/matching-codes`. A detailed description of the mechanisms can be found in Ferrie (1996) and Abramitzky et al. (2014). In this paper, I employ the same code as Bailey et al. (2017a) who kindly provided me with their yet unpublished Stata script (Bailey and Cole, 2018).

[15]See e.g. Atack et al. (1992) for information on phonetic corrections.

procedure of Abramitzky et al. (2012) can be wrapped up as follows: (1—optional) correct names phonetically, (2) search for exact and unique matches with respect to specified characteristics (here: first name, last name, year of birth, place of family origin), (3) if (2) was not successful, search for a match with one year of age difference, and (4) repeat (3) with a bandwidth of two years. The resulting sub-samples are described in Tables 5 and 6.

3 Results

In this section, I present the results of the main analysis that can be split into four parts: (1) geographic mobility (discussed in Section 3.1), (2) life patterns (presented in Section 3.2), (3) linking mechanisms (summarized in Section 3.3), and (4) relative size of bias (analyzed in Section 3.4). In the first three parts, I will shed light on the three potential sources of bias in detail. The fourth part provides insights into how strongly each source of bias affects social mobility estimates compared to the others. Appendix A.1 contains all transition matrices, on which the measures of mobility are based in this section.

3.1 Geographic Mobility

Distribution across Occupational Categories Table 2 reveals differences between migrants and non-migrants. It appears that sons of middle or high SEP (lower white-collar) fathers were more likely to move abroad. This might indicate migration barriers to lower SEP individuals due to limited resources. The migrating sons predominantly entered middle SEP (lower white-collar) occupations. These results regarding the selection of emigrants are similar to the findings for nineteenth century Germany (Wegge, 1999, 2002, 2010). Interestingly, Table 3 suggests large differences of migrants' occupational categories by destination continent.[16] Apparently, sons moving to countries in Europe were more positively selected than those moving to non-European destinations. The fathers of migrants to Europe exhibited the largest share in higher SEP and white-collar occupations across all groups. Similarly, the corresponding sons were even more likely to enter high SEP occupations than the geographically immobile. This picture partially reverts for father-son pairs with sons moving to North America. These individuals appear to have been negatively selected with respect to the SEP they entered but were still positively selected with respect to their fathers' SEPs. Lastly, migrants to South America, Africa, Asia, or Australia appear to have been intermediately selected with respect to SEP of both the fathers and the sons.

---

[16]Table 30 in Appendix A.2 contains details on individuals moving to South America, Africa, Asia, and Australia.

Absolute Mobility I introduce a measure of absolute mobility, the fraction of mobile individuals, to start the analysis of the bias in intergenerational mobility estimates due to migration. Absolute mobility captures the experienced level of social mobility given the occupational distribution across categories. Transition matrices pose as basis for this measure of absolute mobility.[17] They contain the absolute frequency of intergenerational transitions between all possible categories. The fraction of mobile individuals can be calculated by dividing the number of sons of occupational category $i \in \{1, ..., N\}$ fathers that enter a different occupational category $\neg i$, where N denotes the number of categories.[18] If the absolute frequency of sons with i category fathers who enter occupational category j is denoted by X_{ij}, the fraction of socially mobile individuals M is given by

$$M = \frac{\sum_{i=1}^{N} X_{\neg ii}}{\sum_{i=1}^{N} \sum_{j=1}^{N} X_{ij}} = \sum_{i=1}^{N} p_{\neg ii}, \tag{1}$$

where p_{ij} denotes the probability that the son of an occupation i father enters category j.

Figures 1, 2, and 3 display the fraction of mobile father-son pairs by migration status and destination. A distinction between upward (sons move towards higher SEP, or white-collar occupations) and downward (sons move towards lower SEP, or unskilled workers occupations) mobility allows to analyze whether the sons that were socially mobile profited from this mobility or suffered from it. Note that the distinction between upward and downward mobility is more involved with the Long-Ferrie categorizations as they are not meant to be interpreted in an ordinal way. Hence, I only interpret the ratio of upward vs downward mobility employing the SEP categorization. The difference in the share of mobile individuals across migration status is small in all of the applied occupational categorizations. Shifting the focus to the continents of destination exhibits some evidence on differential rates of mobility depending on the country of destination. Especially with respect to the baseline Long-Ferrie categorization, there appears to be a difference between individuals moving to Europe and those moving to North America. Sons relocating

---

[17]Transition matrices also pose as foundation for many measures of relative mobility as presented subsequently.

[18]With both the SEP categorization and the basic Long-Ferrie categorization N is equal to three (low, middle, and high SEP and unskilled workers, skilled workers, and white-collar). In the extended Long-Ferrie categorization with a distinction between higher and lower white-collar occupations N is equal to four.

to European countries exhibited less intergenerational mobility than those that moved to Northern America. The split between upward and downward mobility in Figure 1 displays that geographically immobile sons (Stay) experienced upward mobility 1.3 times more often than downward mobility, whereas this ratio is 0.99 for emigrants. This might indicate that even if the level of absolute mobility only differed marginally between geographically mobile and immobile individuals, there were differences in the structure of social mobility. Apparently, sons migrating to European countries were also positively selected with respect to the chances of upward mobility as compared to sons that preferred North America. The former were 1.12 times more likely to experience upward mobility vs downward mobility while this ratio was 0.63 for the latter.

Relative Mobility   Measures of relative mobility allow to correct for the different sizes of occupational categories in the labor market.[19] Two-way log-odds ratios are one of the easiest methods to quantify relative mobility. Log-odds ratios $\Theta_{2,i}$ quantify the "advantage" sons of category i fathers had to enter the same category vs all other categories over sons of fathers with a different occupational category $\neg i$. They are defined as

$$\Theta_{2,i} = log \left[\frac{p_{ii}/(1 - p_{ii})}{p_{\neg ii}/(1 - p_{\neg ii})} \right].$$ (2)

Two-way log-odds ratios by occupational category and migration status are displayed in Figures 4, 5, and 6. The figures provide further evidence that focusing the analysis of intergenerational mobility on stayers might lead to selection of socially less mobile individuals. For example, geographically immobile sons of high SEP fathers were 6.0 times more likely to enter high SEP occupations as well vs other occupations than sons of middle or low SEP fathers. In the baseline sample including return migrants and emigrants, this number is 4.7. Similarly, emigrating sons were only 3.5 times more likely to follow their father into a high SEP occupation vs other occupations than emigrating sons of low or middle SEP fathers. The Long-Ferrie categorizations produce qualitatively similar but quantitatively less pronounced differences. All three categorizations reveal major differences when splitting geographically mobile sons by destination. This provides

---

[19]See Favre et al. (2018) or Modalsli (2017) for a more detailed introduction into measures of relative mobility.

further evidence on heterogeneous selection of migrants by destination continent—also with respect to the level of relative intergenerational mobility.

The two-way log-odds ratios provide evidence on differences in the level of relative intergenerational mobility by migration status when focusing on the off-diagonal in transition matrices. The Altham (1970b) statistic allows to retrieve a more complete analysis of the transition matrix (see also Altham and Ferrie, 2007; Long and Ferrie, 2013; Modalsli, 2015, 2017; Pérez, 2017).[20] This statistic quantifies the distance of a transition matrix P with dimension N from perfect mobility represented by a matrix of ones J. Following Modalsli (2015, 2017), one can calculate (controlled) Altham statistics by employing multinomial logistic regressions. If one adds control variables (such as age) $\mathbf{X_q}$ when regressing the occupational outcome o_q^s of a son s in the father-son pair q on a set of dummies $\mathbf{D_q} = \{D_1, ..., D_N\}$ indexing the father's occupation, one can estimate the Altham statistic by aggregating the coefficient estimates of the dummies. The set of $N-1$ equations (indexed by k) that have to be estimated can be denoted by

$$log\left[\frac{Pr(o_q^s = k)}{Pr(o_q^s = 1)}\right] = \alpha_k + \beta_k'\mathbf{D_q} + \gamma_k'\mathbf{X_q} + \epsilon_{k,q}, \qquad k = 2, 3, ..., N, \tag{3}$$

where α_k is the estimated constant, γ_k' are the coefficients of the controls, and $\beta_k' = \{\beta_k^1, ..., \beta_k^{N-1}\}$ is the parameter vector of interest. The controlled Altham statistic is then given by

$$d(P, J) = \left[\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{m=1}^{N}\{(\beta_j^i - \beta_m^i) - (\beta_j^l - \beta_m^l)\}^2\right]^{1/2}. \tag{4}$$

Figures 7, 8, and 9 provide the resulting Altham statistics controlled for a quadratic function of both the son's and father's age.[21] The estimates of the Altham statistic support the insights provided by the two-way log-odds ratios: geographically immobile individuals were also socially less mobile. All occupational classifications agree with respect to the direction of the bias whereas they do not regarding the size of bias. The relatively small sample sizes when splitting geographically mobile sons by destination complicates state-

---

[20]The value of the Altham statistic lies between zero and infinity (Altham, 1970b,a). This explains that the imputed confidence intervals are asymmetric in some incidences.

[21]I provide the uncontrolled Altham statistics in Figures 32–34 in Appendix A.2.

ments about significance. However, there are patterns in the point estimates. The point estimates suggest that sons moving to other European countries exhibited a lower level of mobility than e.g. migrants to North America. Overall, the estimates of relative mobility indicate that one might structurally underestimate the level of father-son mobility when excluding emigrating sons from the analysis.

Correlation Coefficient The relation between fathers' and sons' occupations can not only be classified by transition matrices but also by correlation coefficients of cardinal measures. Most studies employing intergenerational correlation coefficients focus on income, education, or elite status outcomes as a basis for the coefficients (see e.g. Black and Devereux, 2011 or Clark, 2014). The cardinal measure employed in this paper is based on occupations once again. I standardize the HISCAM (Lambert et al., 2013) measure associated with each occupation's HISCO code.[22] One can regress the standardized measure of the son $HISCAM_{qs}^{std}$ in the father-son pair q on the corresponding father's standardized measure $HISCAM_{qf}^{std}$ according to

$$HISCAM_{qs}^{std} = \beta HISCAM_{qf}^{std} + \varepsilon_q, \tag{5}$$

which yields the correlation coefficient β.

The correlation coefficient by migration status is depicted in Figure 10. The differences between the baseline sample and the sub-groups by migration status exhibit the same pattern as the previous measures of relative mobility. Emigrating sons show a lower intergenerational correlation indicating a higher level of social mobility than stayers. As with all of the previous measures, there are differences in the level of implied social mobility by destination continent of the son. Irrespective of the destination, all father-son pairs of geographically mobile sons exhibit higher social mobility with respect to the standardized HISCAM than father-son pairs of the baseline sample.[23]

Overall, the analysis of geographic mobility can be boiled down to three observations. First, there was selection of (temporary) international migrants with respect to the oc-

---

[22]The resulting measure has zero mean and a standard deviation of one instead of a range from 0 to 100.

[23]The difference is not statistically significant for sons relocating to South America, Africa, Asia, or Australia.

cupational category of the father and the son. Fathers of geographically mobile sons exhibited higher socioeconomic positions than those of geographically immobile, while these migrating sons tended to attain intermediate socioeconomic positions more often than geographically immobile. Second, this selection translates into different levels of intergenerational mobility. Across most measures, geographically immobile sons exhibited significantly lower levels of intergenerational mobility. Third, there are differences in both selection and the implied level of social mobility across destination continents. Migrants to European countries tended to be better situated than e.g. migrants to North America. Moreover, the former experienced more upward mobility than the latter. The evidence on the direction of differences in overall intergenerational mobility across destinations is inconclusive.

3.2 Life Pattern

Distribution across Occupational Categories Table 4 highlights that of originally 11,384 father-son pairs merely 13 percent remain in the life pattern sample. This means that only 1,476 father-son pairs feature categorizable observations of sons around the age of twenty, thirty, and forty and categorizable observations of fathers around the age of thirty, forty, and fifty. This is caused by timely deaths[24], naturalizations after a certain age[25], and gaps in the observation years. The remaining father-son pairs are positively selected as their SEPs are, on average, higher than in the baseline sample.[26] Table 4 further shows that the fraction of middle SEP, high SEP, and (higher) white-collar individuals increased with age whereas the share of low SEP and (un)skilled individuals decreased with age. This seems natural regarding that career paths usually start at a lower socioeconomic position than they end. Further analyses are required to evaluate whether sample selection with respect to the age is a concern when estimating occupational mobility.

---

[24]The average age at death was fifty-eight in the Zurich data.

[25]If an individual acquires the citizenship of Zurich after the age of e.g. thirty-five, there is no information on occupations at earlier ages, i.e. around thirty.

[26]This is partially caused by lower ages at death of lower SEP individuals. The average age at death for individuals with low, middle, and high SEP as highest occupational outcome was fifty-seven, fifty-nine, and sixty-two, respectively.

Absolute Mobility   I start the analysis of life patterns in social mobility estimates with a measure of absolute mobility. The shares of (upward and downward) mobile individuals according to all classifications are presented in Figures 11, 12, and 13. Overall, there are differences in the estimates on the share of mobile ranging from 42 percent to 46 percent in the SEP categorization and from 31 (51) to 34 (54) percent in the (extended) Long-Ferrie categorization. Nevertheless, the evidence on the direction of the bias due to different ages at categorization of both the father and the son is inconclusive. On average, a higher age at classification of the son induces larger estimates of social mobility in the SEP classification. The Long-Ferrie categorizations do not exhibit such a monotone trend. With respect to the father's age at classification, the Long-Ferrie categorizations suggest (weakly) decreasing mobility whereas the SEP categorization does not feature a monotone trend. Not surprisingly, the prevalence of upward and downward mobility exhibits the same trend across all categorizations. The age of the father negatively (positively) correlates with upward (downward) mobility. The reverse is true for son's age at classification. This finding is predominantly caused by individuals exhibiting upward intragenerational mobility as is depicted in Table 4. If a father is ranked in a "higher" situated occupational category when he is older, the son is less likely to experience upward mobility with respect to that position himself. Similarly, if a sons have, on average, a lower socioeconomic position at lower ages, they are less likely to have already entered an occupation that is "higher" ranked than the one of their fathers. In summary, there are differences in the estimates of absolute intergenerational mobility. Evidence on the direction depending on the age of both the father and the son is mixed.

Relative Mobility   Similar to the results on absolute mobility, the measures of relative mobility do not show a conclusive trend with respect to the age at classification of the father or the son. Even though the two-way log-odds ratios in Figures 14–16 exhibit fluctuations, they do not agree on a clear pattern with respect to categorization ages. The SEP categorization suggests that the classification age of the father affects mobility estimates in a u-shaped way if sons are classified around thirty or forty and in a positive way if sons are classified around twenty. This pattern is not mirrored in the Long-Ferrie categorizations. The Long-Ferrie categorizations do not even exhibit a homogeneous pattern with respect to the father's classification age across occupational categories. If anything,

16

they propose that sons that are categorized at younger ages exhibit lower levels of occupational mobility. Apart from these observations, there is no monotone or homogeneous pattern observable with respect to the ages at classification.

The Altham statistics do not exhibit strong trends either (see Figures 17–19).[27] The Altham statistic of the SEP categorization suggests that mobility was lowest if the sons are classified around 40 and highest if the sons are classified around 20. The point estimates diverge more with higher ages at classification of the father. However, these differences are statistically insignificant. The Long-Ferrie categorizations indicate only a small effect of the age at classification of both the father and the son. All in all, measures of relative mobility indicate that the age at classification may affect the point estimate of intergenerational mobility estimates, but they do not exhibit a clear trend in the direction of the deviations.

Correlation Coefficient   Interestingly, the correlation coefficient of the standardized HISCAM measure depicts (insignificant) trends in both father's and son's age at classification. Figure 20 hints at a negative correlation of intergenerational mobility with both the father's and the son's categorization age.[28] There are several potential explanations for these patterns. First, occupations at later ages might be more representative of the lifetime occupational potential. This explanation would imply that life-cycle bias might, in fact, be a concern when estimating occupational mobility with the HISCAM correlation coefficient. Second, the older the father at the age of classification the closer the son was to an actual occupational choice. Whether this choice was made by him (positively influenced by father) or arranged by the father could not be investigated with the data at hand. Third, the older the son the more likely the father was to be deceased. Consequently, especially sons of self-employed fathers might have inherited the father's business. All of these hypotheses are possible explanations for the observed trends but cannot be tested with the data at hand. Furthermore, the correlation coefficient is the only measure of occupational mobility exhibiting such clear and monotone trends with respect to the classification ages. Hence, the evidence on the direction of bias due to life-cycle patterns

---

[27]Estimates of the uncontrolled Altham statistics are presented in Figures 35–37 in Appendix A.2

[28]The trend in the son's age at classification mirrors the trend suggested by the SEP-based Altham statistic.

is inconclusive.

3.3 Linking Procedures

Performance and Distribution across Occupational Categories   In this section, I present the results from applying automated linking procedures to the artificially separated father-son data. Table 7 contains an evaluation of the linking procedures. The match rates are very high at between 77 and 95 percent correct matches. Similarly, the share of type I errors (wrongly linked father-son pairs) is negligible with values below three per mille. The procedure of Ferrie (1996) produces marginally higher rates of type I errors and does not exhibit strong differences across phonetic name cleaning methods. The highest match rate both with respect to total and correct matches and lowest rate of type I error is achieved with the procedure by Abramitzky et al. (2012, 2014) with Soundex name cleaning. Not surprisingly, Bailey et al. (2017a) find much higher error rates. Their estimates for the match rate (1 – type II error rate) lie between 20 and 40 percent for the same procedures. Similarly, they find a share of type I errors (false positives) between 22 and 43 percent. This depicts nicely that the sample I employ in this paper is not representative as the pool for potential matches is small and easily separable with respect to the linking characteristics (name, year of birth, place of origin).[29] Consequently, the conclusions from this paper can only be extended to the performance of linking mechanisms in data with similar quality and quantity.

Tables 5 and 6 show that the high match rates and small fraction of false links translate into small differences with respect to the distribution across occupational categories and average age. Apparently, middle SEP and (lower) white-collar individuals were more likely to be matched both among fathers and sons. This skews the occupational distribution marginally in that direction. Similarly, younger individuals were matched more frequently. Nevertheless, the deviations are minimal (especially compared to the results from Bailey et al., 2017a). Based on these findings, one would not expect that any of the intergenerational mobility estimates is significantly biased in the linked samples.

---

[29]In other words, there have to be two individuals in the pool of potential son matches with very similar names, year of birth, and exact same place of family origin to induce type I and II errors. As there is no misspelling in names nor errors in the year of birth, the automated linking procedures perform neatly.

Estimates of Intergenerational Mobility   As the solid performance of all automated linking procedures suggests, none of the measures of absolute and relative mobility exhibit large nor significant bias as compared to the baseline sample. Interestingly, employing Ferrie (1996) without name cleaning gets closest to the baseline sample with respect to mobility estimates even though it is outperformed with respect to match rates and false matches by Abramitzky et al. (2012, 2014) with Soundex name cleaning. Figures 21, 22, and 23 show that the share of mobile individuals is stable across mechanisms. There are minor differences with respect to the prevalence of upward vs downward mobility. The linked samples are somewhat more likely to contain upward mobile father-son pairs.

The two-way log-odds ratios depicted in Figures 24–26 point out minor differences too. Most linking mechanisms estimate higher log-odds ratios for the groups of high SEP and unskilled workers, whereas they estimate lower odds ratios for skilled workers and the lower white-collar group. Again, these differences are small. For example, the baseline estimate suggests that sons of high SEP fathers were 5.20 times more likely to enter the same occupational category vs another category than sons of low or middle SEP fathers. The corresponding estimate in the Abramitzky (NYSIIS) sample lies close at 5.45. Aggregating all log-odds ratios into the Altham statistic does not change the picture. Figures 27–29 depict that the point estimates vary somewhat but are always close to the baseline estimate.[30] The occupational categorizations do not agree with respect to the direction of the bias based on the point estimates. The SEP categorization and the extended Long-Ferrie classification imply that linked samples marginally underestimate mobility whereas the three-category Long-Ferrie classification points towards overestimation of mobility in linked samples. Finally, the HISCAM correlation coefficient (Figure 30) further solidifies the impression that the employed linking procedures do not lead to significantly biased estimates of intergenerational mobility in the data at hand as the point estimates are virtually the same across all samples.

3.4   Relative Size of Bias

In this section, I evaluate the relative size of bias due to the three different sources. The previous sections have already lined out that one should expect migration to have the

---

[30]The uncontrolled Altham statistics are presented in Appendix A.2 (Figures 38–40).

largest or at least most consistent impact. The expectation on the impact of different classification ages is unclear whereas automated linking procedures should not exhibit relatively large deviations. Tables 8, 9, and 10 display the direction of the bias due to each source and the relative size in percent of the baseline sample's estimate for all of the employed measures of intergenerational mobility. Figure 31 depicts the results for the correlation coefficient based on HISCAM to provide a representative graphical illustration. In order to boil down the results from the previous sections, I only present the bias in selected sub-samples. Namely, I compare all estimates of six samples: (1) the baseline sample including all father-son pairs, (2) the sample with geographically immobile sons[31], (3) the linked sample employing the procedure of Ferrie (1996) without name cleaning, and (4)–(6) three combinations of father's and son's age at classification (son at twenty and father at fifty (4), both at forty (5), and son at forty and father at thirty (6)). Note that the estimated level of mobility in the last three sub-samples are not as easily comparable to the baseline sample as the former two. Of course, one can see which of the selected ages produce estimates closest to the baseline sample but I excluded all father-son pairs that were not observable at all ages in these sub-samples. Consequently, comparing the three with each other gives better insights in the relative size of the bias.

The results suggest that restricting the sample to geographically immobile individuals consistently biases the estimate of intergenerational mobility downwards by between one and ten percent. The bias due to employing linking procedures is comparably negligible. It is between four and eighty-five times smaller than the migration bias and never exceeds one percent of the baseline estimate. The life pattern estimates differ substantially from the baseline sample as well. These differences between the life pattern samples and the baseline sample are both due to sample selection (excluding all individuals not observable at every age) and differences in the level of social mobility caused by differences in age. Consequently, the comparison with the other sources of bias is to be taken with a grain of salt. Comparing the three life pattern samples among each other seems more appropriate. This comparison reveals that the deviations between the three samples' estimates usually range up to ten percent of the baseline sample. This suggests that the relative size of bias due to life patterns may be roughly comparable to the relative size of migration bias.

---

[31]This poses as direct comparison to the most prevalent scenario in the existing literature on social mobility, as most analyses are restricted to the non-migrating population.

However, there is no apparent ordering of the three classification age combinations across different occupational categorizations or measures of mobility.

All in all, the results on bias due to each of the three sources can be summarized as follows. Firstly, narrowing the analysis down to geographically immobile individuals underestimates the level of intergenerational mobility with deviations between one and ten percent (depending on the measure of mobility). Secondly, life patterns or the age at classification affect the estimated level of occupational mobility but the results are inconclusive with respect to trends, direction, and relative size of the bias. Thirdly, automated linking procedures do not generate significant bias in social mobility estimates in the data base of Zurich's male citizenry in the nineteenth century.

4 Conclusion

This paper contributes to several strands of the literature related to the topic of intergenerational mobility such as migration, life-cycle bias, and record linkage by employing data on Zurich's citizenry between 1799 and 1926. The data are unique because they contain observable intergenerational links, and allow to track individuals over the course of their lives and to follow them after emigration. These features enable me to evaluate potential biases in estimates of social mobility. I explore the direction and size of distortions due to the following sources: (1) migration, (2) life-cycle patterns in occupational outcomes, and (3) record linkage. Each bias can be set into relation in order to highlight where future research has to be particularly careful. In addition, the analysis is based on a broad set of measures for absolute and relative occupational mobility.

The results can be boiled down to four main findings. First, emigrants were intermediately selected with differences by country of destination. Zurich emigrants to European countries were more positively selected as compared to emigrants to the United States and Canada in the nineteenth century. Second, Zurich citizens exhibited an occupational life pattern indicating that individuals experienced non-negligible levels of intragenerational (upward) mobility. The older a male citizen was the higher his socioeconomic position. Third, state-of-the-art record linkage procedures perform well due to the detail of information and the small size of the populations to match. On average, around eighty-five percent of father-son pairs could be matched and less than three per mille of matches were wrongly assigned. Fourth, excluding emigrating individuals underestimates the level of social mobility of all father-son pairs by an average of four percent. Life patterns in the occupational distribution affect estimates of intergenerational mobility on a comparable scale but do not exhibit a monotone pattern with respect to age of the father or the son. Due to their neat performance, linking procedures do not induce social mobility estimates to deviate significantly from "true" estimates in non-linked data. Consequently, future research should aim at addressing all of the raised issues depending on the quality and number of the data at hand.

References

ABOWD, J. M. (2017): "Large-scale Data Linkage from Multiple Sources: Methodology and Research Challenges," Tech. rep., NBER Summer Institute Methods Lecture.

ABRAMITZKY, R. AND L. BOUSTAN (2017): "Immigration in American economic history," *Journal of economic literature*, 55, 1311–45.

ABRAMITZKY, R., L. P. BOUSTAN, AND K. ERIKSSON (2012): "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration," *The American Economic Review*, 102, 1832–1856.

——— (2013): "Have the poor always been less likely to migrate? Evidence from inheritance practices during the age of mass migration," *Journal of Development Economics*, 102, 2–14.

——— (2014): "A nation of immigrants: Assimilation and economic outcomes in the age of mass migration," *Journal of Political Economy*, 122, 467–506.

ABRAMITZKY, R., R. MILL, AND S. PÉREZ (2019): "Linking individuals across historical sources: a fully automated approach," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 1–18.

ALTHAM, P. M. (1970a): "The measurement of association in a contingency table: three extensions of the cross-ratios and metrics methods," *Journal of the Royal Statistical Society. Series B (Methodological)*, 32 (3), 395–407.

——— (1970b): "The Measurement of Association of Rows and Columns for an r× s Contingency Table," *Journal of the Royal Statistical Society. Series B (Methodological)*, 32 (1), 63–73.

ALTHAM, P. M. AND J. P. FERRIE (2007): "Comparing Contingency Tables. Tools for Analyzing Data from Two Groups Cross-Classified by Two Characteristics," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 40, 3–16.

ATACK, J., F. BATEMAN, AND M. E. GREGSON (1992): ""Matchmaker, Matchmaker, Make Me a Match" A General Personal Computer-Based Matching Program for His-

torical Research," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 25, 53–65.

Bailey, M., C. Cole, M. Henderson, and C. Massey (2017a): "How well do automated linking methods perform in historical samples? Evidence from new ground truth," NBER Working Papers 24019, National Bureau of Economic Research.

Bailey, M., C. Cole, and C. Massey (2017b): "Representativeness and False Links in the 1850-1930 IPUMS Linked Representative Historical Samples," Tech. rep., University of Michigan National Bureau of Economic Research.

Bailey, M. J. and C. Cole (2018): "autolink.ado," Unpublished stata ado-file, University of Michigan, Department of Economics.

Barone, G. and S. Mocetti (2016): "Intergenerational Mobility in the Very Long Run: Florence 1427-2011," Working Paper 1060, Bank of Italy, https://ssrn.com/abstract=2856359.

Baskerville, P., L. Dillon, K. Inwood, E. Roberts, S. Ruggles, K. Schürer, and J. R. Warren (2014): "Mining Microdata: Economic Opportunity and Spatial Mobility in Britain and the United States, 1850-1881," in *2014 IEEE International Conference on Big Data (Big Data)*, 5–13.

Bauer, P. and R. T. Riphahn (2007): "Heterogeneity in the Intergenerational Transmission of Educational Attainment: Evidence from Switzerland on Natives and Second-Generation Immigrants," *Journal of Population Economics*, 20, 121–148.

Black, S. and P. Devereux (2011): "Recent Developments in Intergenerational Mobility," in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card, Elsevier, vol. 4B, chap. 16, 1487–1541, 1 ed.

Böhlmark, A. and M. J. Lindquist (2006): "Life-Cycle Variations in the Association between Current and Lifetime Income: Replication and Extension for Sweden," *Journal of Labor Economics*, 24, 879–896.

Borjas, G. J. (1992): "Ethnic Capital and Intergenerational Mobility," *The Quarterly Journal of Economics*, 107, 123–150.

———— (1993): "The Intergenerational Mobility of Immigrants," *Journal of Labor Economics*, 11, 113–135.

———— (1994): "Long-run Convergence of Ethnic Skill Differentials: The Children and Grandchildren of the Great Migration," *ILR Review*, 47, 553–573.

———— (1995): "Ethnicity, Neighborhoods, and Human-Capital Externalities," *The American Economic Review*, 85, 365–390.

BOURDIEU, J., J. P. FERRIE, AND L. KESZTENBAUM (2009): "Vive la différence? Intergenerational mobility in France and the United States during the nineteenth and twentieth centuries," *Journal of Interdisciplinary History*, 39, 523–557.

CARD, D. (2005): "Is the New Immigration Really so Bad?" *The Economic Journal*, 115, 300–323.

CHRISTEN, P. AND K. GOISER (2007): "Quality and Complexity Measures for Data Linkage and Deduplication." *Quality Measures in Data Mining*, 43, 127–151.

CLARK, G. (2014): *The Son also Rises: Surnames and the History of Social Mobility*, Princeton and Oxford: Princeton University Press.

COLLINS, W. J. AND M. H. WANAMAKER (2017): "Up from Slavery? African American Intergenerational Economic Mobility Since 1880," Tech. rep., National Bureau of Economic Research.

CONSTANT, A. F. AND K. F. ZIMMERMANN (2013): *International handbook on the economics of migration*, Edward Elgar Publishing.

DRIBE, M., J. HELGERTZ, AND B. VAN DE PUTTE (2015): "Did social mobility increase during the industrialization process? A micro-level study of a transforming community in southern Sweden 1828-1968," *Research in Social Stratification and Mobility*, 41, 25–39.

ERIKSSON, B. (2017): "False Positves and Faulty Estimates: Linked Census Data and Bias to Estimates of Social Mobility." Presentation at The Systematic Linking of Historical Records Workshop, University of Guelph, Centre for Economic Demography - Lund University.

FAVRE, G., J. FLORIS, AND U. WOITEK (2018): "Intergenerational Mobility in the 19th Century: Micro-Level Evidence from the City of Zurich," ECON - Working Papers 274, Department of Economics - University of Zurich.

FEIGENBAUM, J. J. (2018): "Multiple Measures of Historical Intergenerational Mobility: Iowa 1915 to 1940," *The Economic Journal*, 128, F446–F481.

FERRIE, J. P. (1996): "A New Sample of Males Linked from the Public Use Microdata Sample of the 1850 U.S. Federal Census of Population to the 1860 U.S. Federal Census Manuscript Schedules," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 29, 141–156.

——— (2004): "Longitudinal Data for the Analysis of Mobility in the U.S., 1850-1910," Working paper, Department of Economics and Institute for Policy Research, Northwestern University.

——— (2005): "History Lessons: The End of American Exceptionalism? Mobility in the United States Since 1850," *Journal of Economic Perspectives*, 19, 199–215.

GOEKEN, R., L. HUYNH, T. LYNCH, AND R. VICK (2011): "New Methods of Census Record Linking," *Historical Methods*, 44, 7–14.

GRAWE, N. D. (2006): "Lifecycle Bias in Estimates of Intergenerational Earnings Persistence," *Labour Economics*, 13, 551–570.

GREGG, P., L. MACMILLAN, AND C. VITTORI (2017): "Moving Towards Estimating Sons' Lifetime Intergenerational Economic Mobility in the UK," *Oxford Bulletin of Economics and Statistics*, 79, 79–100.

HAIDER, S. AND G. SOLON (2006): "Life-Cycle Variation in the Association between Current and Lifetime Earnings," *The American Economic Review*, 96, 1308–1320.

HAMMARSTEDT, M. AND M. PALME (2006): "Intergenerational Mobility, Human Capital Transmission and the Earnings of Second-Generation Immigrants in Sweden," IZA Discussion Papers 1943, Institute for the Study of Labor (IZA).

HERZOG, T. N., F. J. SCHEUREN, AND W. E. WINKLER (2007): *Data Quality and Record Linkage Techniques*, New York: Springer Science & Business Media.

JENKINS, S. (1987): "Snapshots versus Movies:"Lifecycle Biases" and the Estimation of Intergenerational Earnings Inheritance," *European Economic Review*, 31, 1149–1158.

LAMBERT, P. S., R. L. ZIJDEMAN, M. H. VAN LEEUWEN, I. MAAS, AND K. PRANDY (2013): "The construction of HISCAM: A stratification scale based on social interactions for historical comparative research," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 46, 77–89.

LONG, J. (2013): "The surprising social mobility of Victorian Britain," *European Review of Economic History*, 17, 1–23.

LONG, J. AND J. P. FERRIE (2013): "Intergenerational Occupational Mobility in Great Britain and the United States since 1850," *The American Economic Review*, 103, 1109–1137.

MASSEY, C. G. (2017): "Playing with Matches: An Assessment of Accuracy in Linked Historical Data," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 1–15.

MODALSLI, J. (2015): "Estimating occupational mobility with covariates," *Economics Letters*, 133, 77–80.

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

NYBOM, M. AND J. STUHLER (2016a): "Biases in Standard Measures of Intergenerational Income Dependence," *Journal of Human Resources*, 0715–7290R.

——— (2016b): "Heterogeneous Income Profiles and Lifecycle Bias in Intergenerational Mobility Estimation," *Journal of Human Resources*, 51, 239–268.

PÉREZ, S. (2017): "Intergenerational Occupational Mobility Across Three Continents: Were the Americas Exceptional?" Unpublished manuscript, Department of Economics, University of California, Davis.

RUGGLES, S., C. A. FITCH, AND E. ROBERTS (2018): "Historical census record linkage," *Annual review of sociology*, 44, 19–37.

SCHEUREN, F. AND W. WINKLER (1993): "Regression Analysis of Data Files that are Computer Matched," *Survey Methodology*, 19, 39–58.

SCHÜREN, R. (1989): *Soziale Mobilität. Muster, Veränderungen und Bedingungen im 19. Und 20. Jahrhundert*, St. Katharinen: Scripta Mercaturae Verlag.

SOLON, G. (1999): "Intergenerational Mobility in the Labor Market," in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card, Elsevier, vol. 3, Part A, chap. 29, 1761–1800, 1 ed.

VAN LEEUWEN, M. H. AND I. MAAS (2011): *HISCLASS: A Historical International Social Class Scheme*, Universitaire Pers Leuven.

VAN LEEUWEN, M. H., I. MAAS, AND A. MILES (2002): *HISCO: Historical International Standard Classification of Occupations*, Leuven Univ Pr.

WARD, Z. (2017): "The Not-So-Hot Melting Pot: The Persistence of Outcomes for Descendants of the Age of Mass Migration," Unpublished manuscript, Australian National University.

WEGGE, S. A. (1999): "To part or not to part: emigration and inheritance institutions in nineteenth-century Hesse–Cassel," *Explorations in economic history*, 36, 30–55.

——— (2002): "Occupational self-selection of European emigrants: Evidence from nineteenth-century Hesse-Cassel," *European Review of Economic History*, 6, 365–394.

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

Figures and Tables

FIGURE 1. SHARE OF (UPWARD AND DOWNWARD) MOBILE ACCORDING TO SEP CATEGORIES BY MIGRATION STATUS.
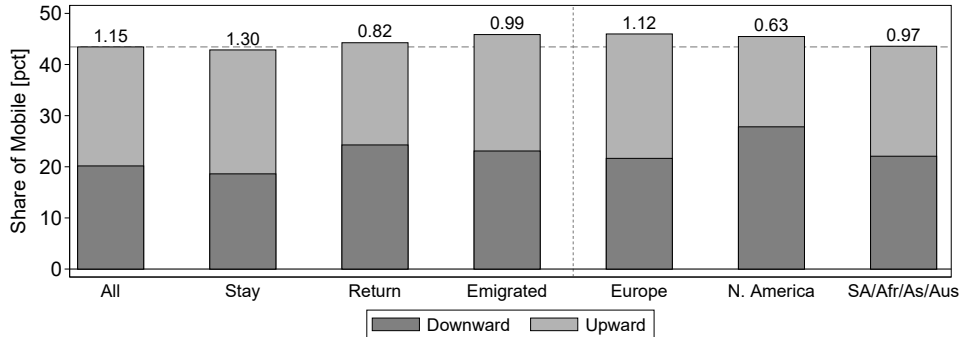
Note: The height of the bars displays the share of mobile sons that did not enter the same occupational category as their father (off-diagonal in the transition matrices). Upward mobile individuals enter higher SEP than their father. The numbers above the bars denote the ratio of upward to downward mobility. The dashed horizontal line marks the share of mobile in the baseline sample. The groups to on the RHS of the vertical dashed line are geographically mobile (Return and Emigrated) split by destination.

FIGURE 2. SHARE OF (UPWARD AND DOWNWARD) MOBILE ACCORDING TO LONG-FERRIE CATEGORIES BY MIGRATION STATUS.

Note: The height of the bars displays the share of mobile sons that did not enter the same occupational category as their father (off-diagonal in the transition matrices). Upward mobility denotes individuals that move closer to white-collar occupations. The numbers above the bars denote the ratio of upward to downward mobility. The dashed horizontal line marks the share of mobile in the baseline sample. The groups to on the RHS of the vertical dashed line are geographically mobile (Return and Emigrated) split by destination.
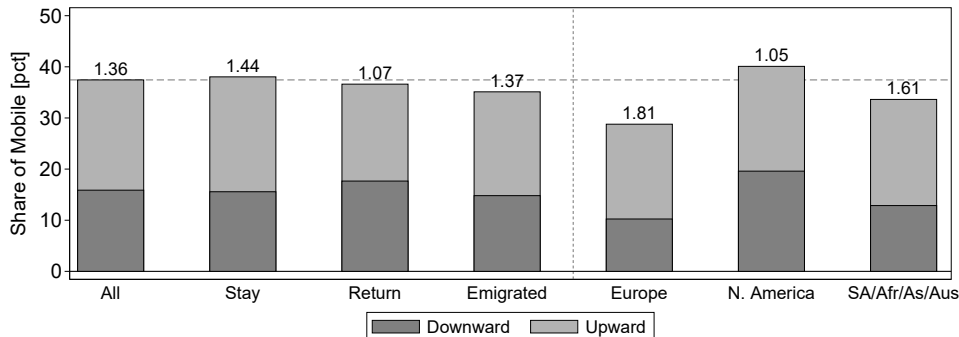
FIGURE 3. SHARE OF (UPWARD AND DOWNWARD) MOBILE ACCORDING TO EXTENDED LONG-FERRIE CATEGORIES BY MIGRATION STATUS.

Note: The height of the bars displays the share of mobile sons that did not enter the same occupational category as their father (off-diagonal in the transition matrices). Upward mobility denotes individuals that move closer to higher white-collar occupations. The numbers above the bars denote the ratio of upward to downward mobility. The dashed horizontal line marks the share of mobile in the baseline sample. The groups to on the RHS of the vertical dashed line are geographically mobile (Return and Emigrated) split by destination.

FIGURE 4. TWO-WAY LOG-ODDS RATIOS ACCORDING TO SEP CATEGORIES BY MIGRATION STATUS.
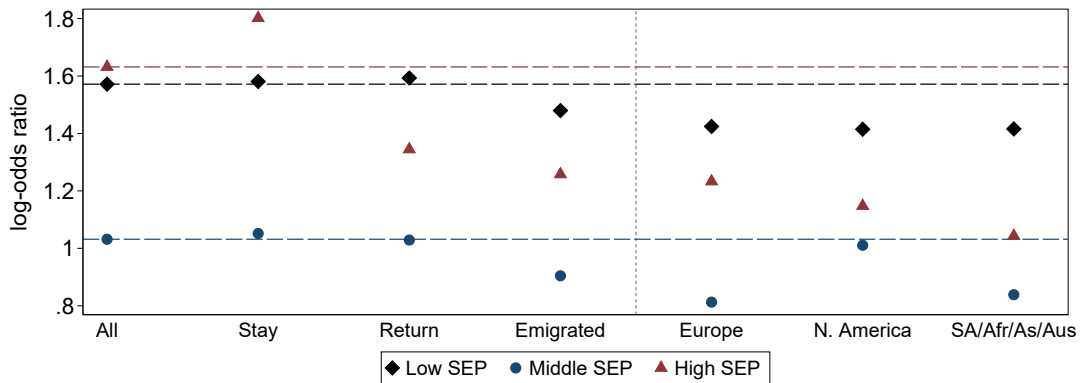
Note: The dashed lines denote the level of the log-odds ratios in the baseline sample (color-coded). The groups to on the RHS of the vertical dashed line are geographically mobile (Return and Emigrated) split by destination.

FIGURE 5. TWO-WAY LOG-ODDS RATIOS ACCORDING TO LONG-FERRIE CATEGORIES BY MIGRATION STATUS.

Note: The dashed lines denote the level of the log-odds ratios in the baseline sample (color-coded). The groups to on the RHS of the vertical dashed line are geographically mobile (Return and Emigrated) split by destination.

FIGURE 6. TWO-WAY LOG-ODDS RATIOS ACCORDING TO EXTENDED LONG-FERRIE CATEGORIES BY MIGRATION STATUS.
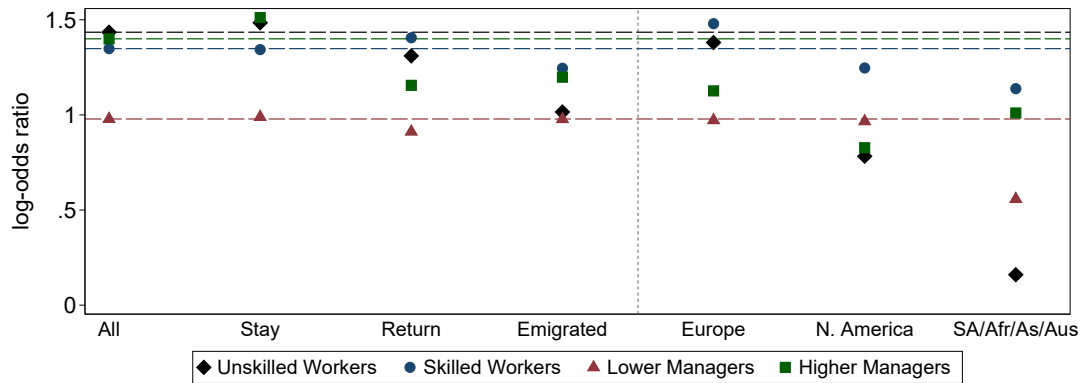
Note: The dashed lines denote the level of the log-odds ratios in the baseline sample (color-coded). The groups to on the RHS of the vertical dashed line are geographically mobile (Return and Emigrated) split by destination.

FIGURE 7. ALTHAM STATISTIC ACCORDING TO SEP CATEGORIES BY MIGRATION STATUS.
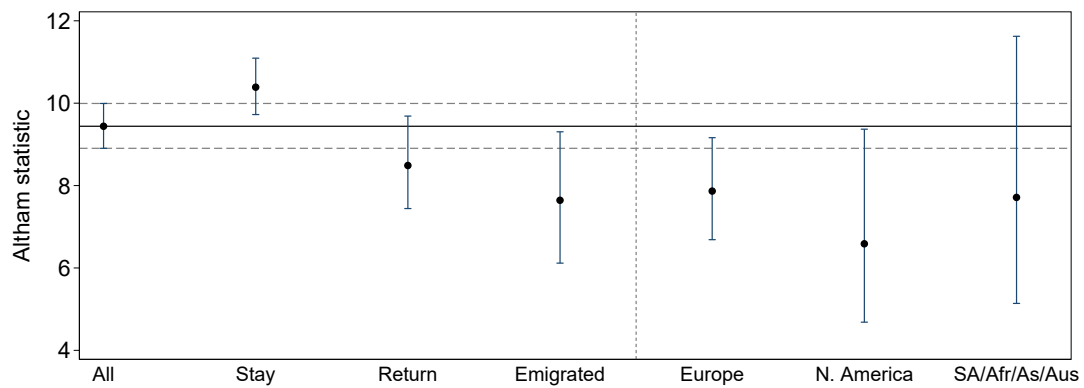
Note: This figure contains the Altham statistic controlled for a quadratic function of the father's and the son's age. The lines mark the level (solid) and confidence intervals (dashed) for the baseline sample. Confidence intervals are calculated by a bootstrapping procedure as explained in Modalsli (2015, p. 8). The groups to on the RHS of the vertical dashed line are geographically mobile (Return and Emigrated) split by destination.

FIGURE 8. ALTHAM STATISTIC ACCORDING TO LONG-FERRIE CATEGORIES BY MIGRATION STATUS.
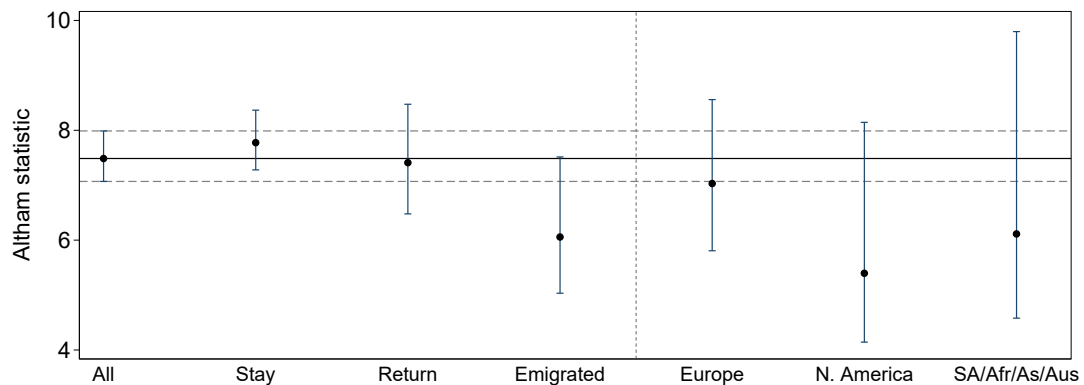
Note: This figure contains the Altham statistic controlled for a quadratic function of the father's and the son's age. The lines mark the level (solid) and confidence intervals (dashed) for the baseline sample. Confidence intervals are calculated by a bootstrapping procedure as explained in Modalsli (2015, p. 8). The groups to on the RHS of the vertical dashed line are geographically mobile (Return and Emigrated) split by destination.

FIGURE 9. ALTHAM STATISTIC ACCORDING TO EXTENDED LONG-FERRIE CATEGORIES BY MIGRATION STATUS.
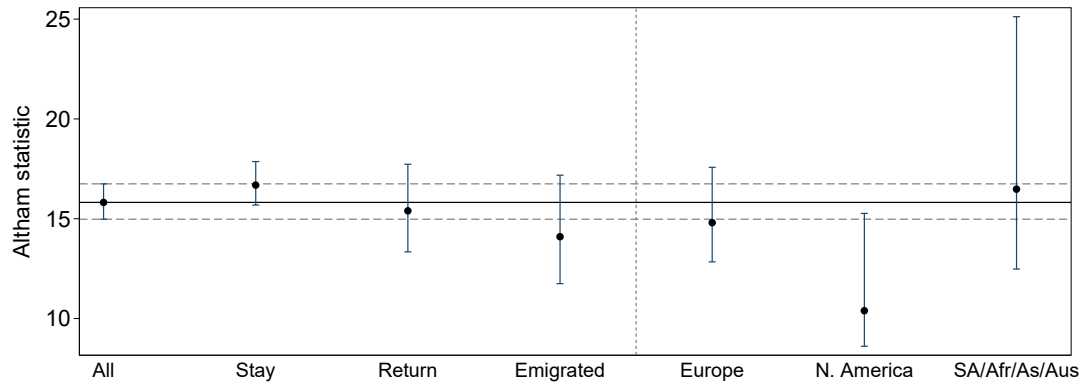
Note: This figure contains the Altham statistic controlled for a quadratic function of the father's and the son's age. The lines mark the level (solid) and confidence intervals (dashed) for the baseline sample. Confidence intervals are calculated by a bootstrapping procedure as explained in Modalsli (2015, p. 8). The groups to on the RHS of the vertical dashed line are geographically mobile (Return and Emigrated) split by destination.

FIGURE 10. CORRELATION COEFFICIENT OF THE STANDARDIZED HISCAM MEASURE BY MIGRATION STATUS.
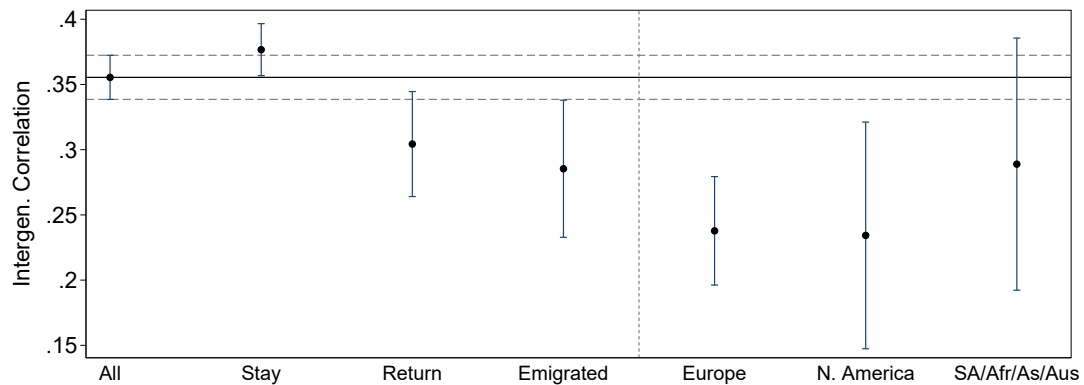
Note: The solid (dashed) black line represents the estimate (confidence intervals) for the baseline sample. The groups to on the RHS of the vertical dashed line are geographically mobile (Return and Emigrated) split by destination.

FIGURE 11. SHARE OF (UPWARD AND DOWNWARD) MOBILE ACCORDING TO SEP CATEGORIES BY AGE AT CLASSIFICATION.

Note: The age at classification of the son is color coded, the father's age at classification is on the x-axis. The share of mobile denotes sons that did not enter the same occupational category as their father (off-diagonal in the transition matrices). Upward mobile individuals enter higher SEP than their father. The horizontal line marks the share of mobile of the 40-40 sample (both son and father categorized at age 40).

FIGURE 12. SHARE OF (UPWARD AND DOWNWARD) MOBILE ACCORDING TO LONG-FERRIE CATEGORIES BY AGE AT CLASSIFICATION.

Note: The age at classification of the son is color coded, the father's age at classification is on the x-axis. The share of mobile denotes sons that did not enter the same occupational category as their father (off-diagonal in the transition matrices). Upward mobility denotes individuals that move closer to white-collar occupations. The horizontal line marks the share of mobile of the 40-40 sample (both son and father categorized at age 40).

FIGURE 13. SHARE OF (UPWARD AND DOWNWARD) MOBILE ACCORDING TO EXTENDED LONG-FERRIE CATEGORIES BY AGE AT CLASSIFICATION.

Note: The age at classification of the son is color coded, the father's age at classification is on the x-axis. The share of mobile denotes sons that did not enter the same occupational category as their father (off-diagonal in the transition matrices). Upward mobility denotes individuals that move closer to higher white-collar occupations. The horizontal line marks the share of mobile of the 40-40 sample (both son and father categorized at age 40).

37

FIGURE 14. TWO-WAY LOG-ODDS RATIOS ACCORDING TO SEP CATEGORIES BY AGE AT CLASSIFICATION.

Note: The age at classification of the son is color coded, the father's age at classification is on the x-axis. The dashed lines denote the level of the log-odds ratios of the 40-40 sample (both son and father categorized at age 40).

FIGURE 15. TWO-WAY LOG-ODDS RATIOS ACCORDING TO LONG-FERRIE CATEGORIES BY AGE AT CLASSIFICATION.

Note: The age at classification of the son is color coded, the father's age at classification is on the x-axis. The dashed lines denote the level of the log-odds ratios of the 40-40 sample (both son and father categorized at age 40).

FIGURE 16. TWO-WAY LOG-ODDS RATIOS ACCORDING TO EXTENDED LONG-FERRIE CATEGORIES BY AGE AT CLASSIFICATION.
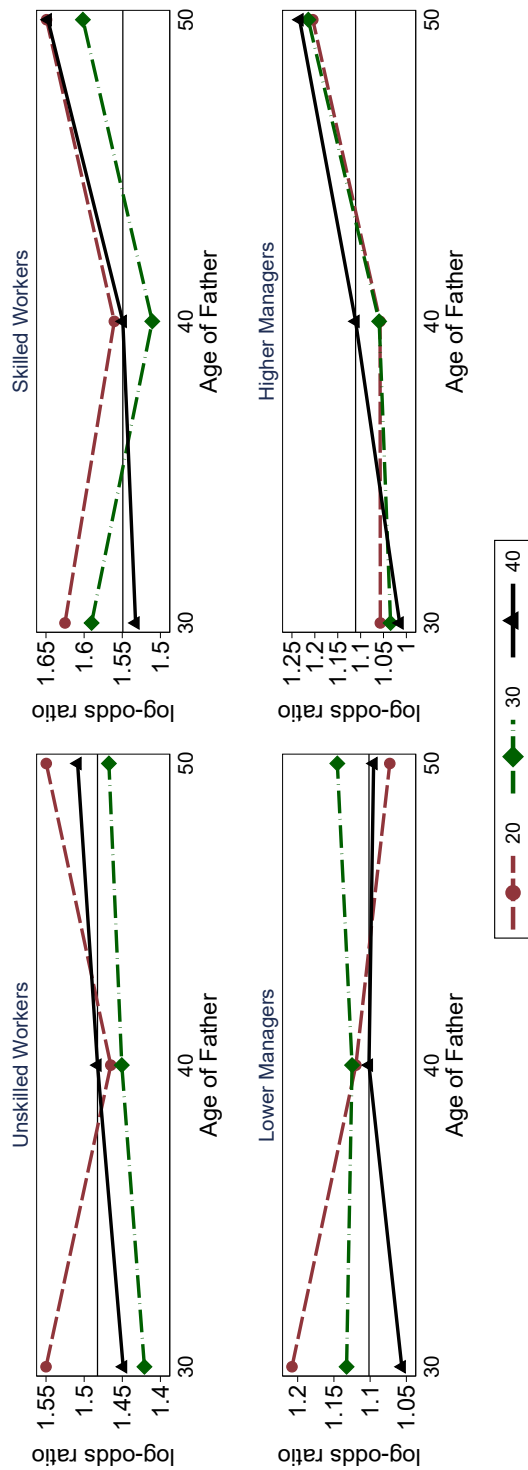
Note: The age at classification of the son is color coded, the father's age at classification is on the x-axis. The dashed lines denote the level of the log-odds ratios of the 40-40 sample (both son and father categorized at age 40).

FIGURE 17. ALTHAM STATISTIC ACCORDING TO SEP CATEGORIES BY AGE AT CLASSIFICATION.
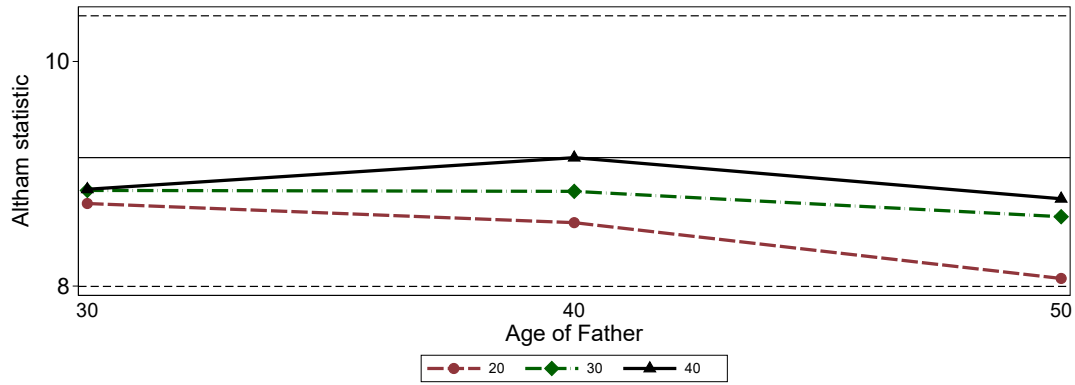
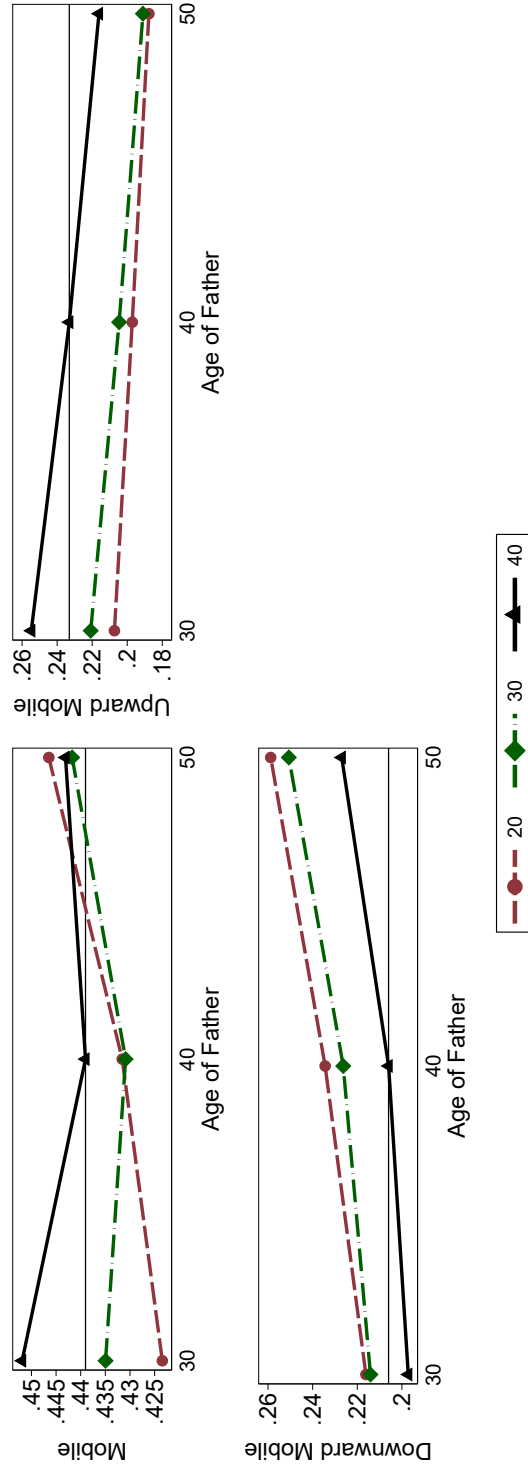*Note*: The age at classification of the son is color coded, the father's age at classification is on the x-axis. This figure contains the Altham statistic controlled for a quadratic function of the father's and the son's age. The lines mark the level (solid) and confidence intervals (dashed) of the 40-40 sample (both son and father categorized at age 40). Confidence intervals are calculated by a bootstrapping procedure as explained in Modalsli (2015, p. 8).

FIGURE 18. ALTHAM STATISTIC ACCORDING TO LONG-FERRIE CATEGORIES BY AGE AT CLASSIFICATION.
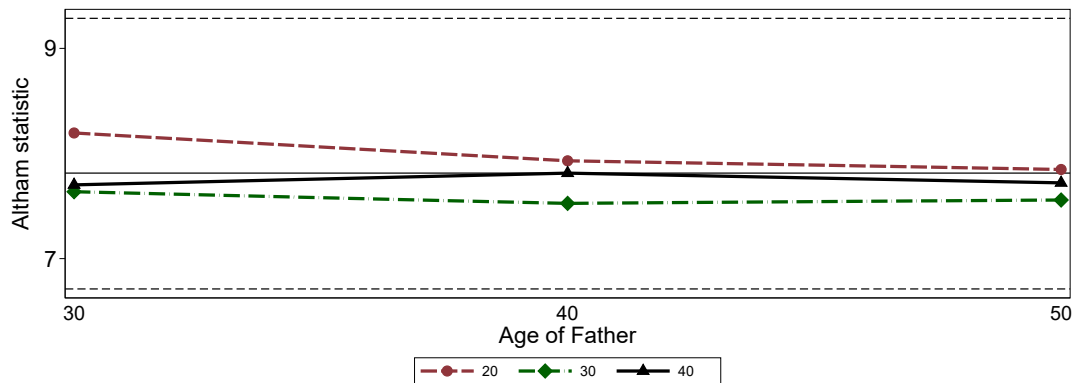
Note: The age at classification of the son is color coded, the father's age at classification is on the x-axis. This figure contains the Altham statistic controlled for a quadratic function of the father's and the son's age. The lines mark the level (solid) and confidence intervals (dashed) of the 40-40 sample (both son and father categorized at age 40). Confidence intervals are calculated by a bootstrapping procedure as explained in Modalsli (2015, p. 8).

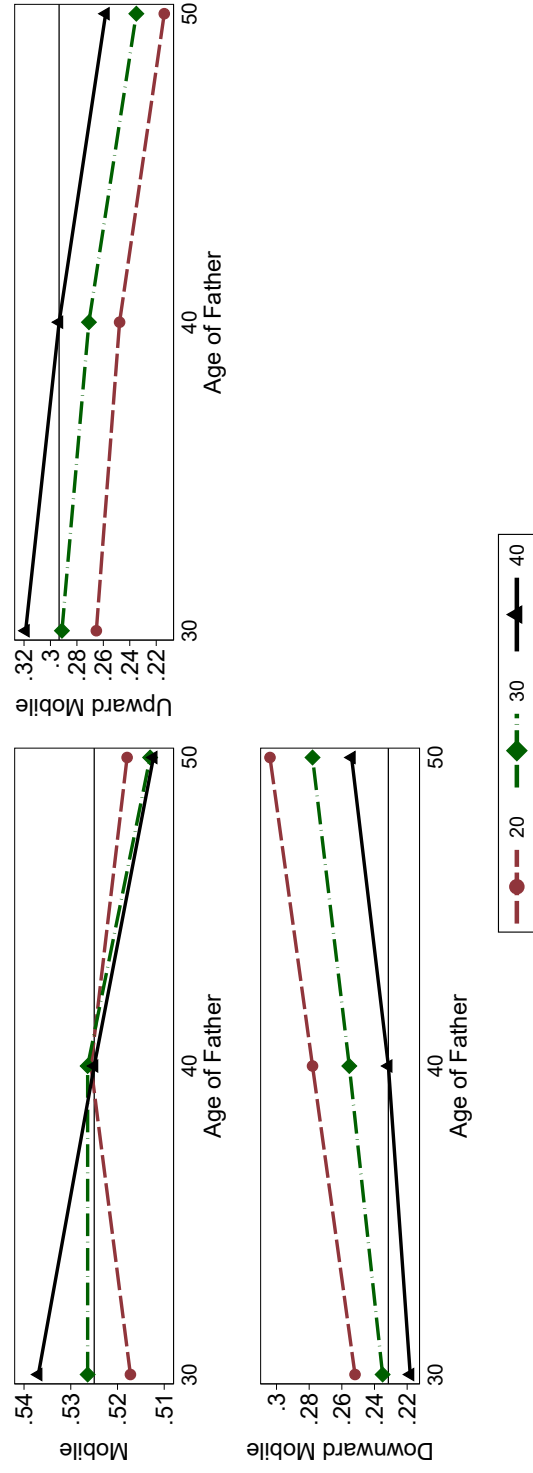FIGURE 19. ALTHAM STATISTIC ACCORDING TO EXTENDED LONG-FERRIE CATEGORIES BY AGE AT CLASSIFICATION.

Note: The age at classification of the son is color coded, the father's age at classification is on the x-axis. This figure contains the Altham statistic controlled for a quadratic function of the father's and the son's age. The lines mark the level (solid) and confidence intervals (dashed) of the 40-40 sample (both son and father categorized at age 40). Confidence intervals are calculated by a bootstrapping procedure as explained in Modalsli (2015, p. 8).

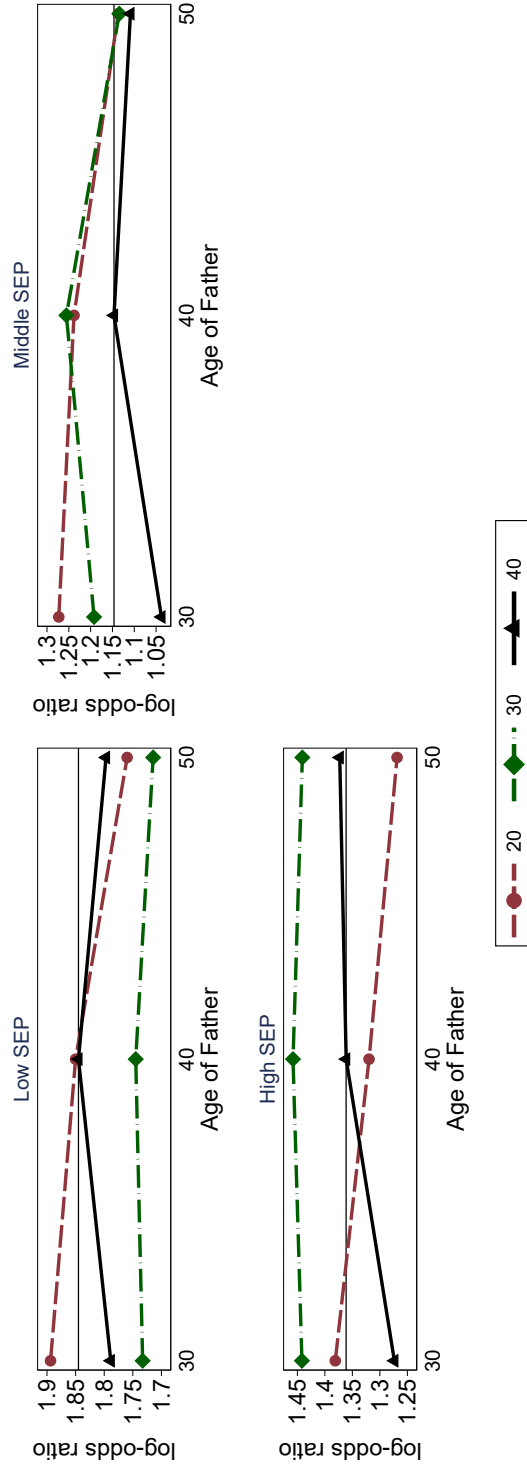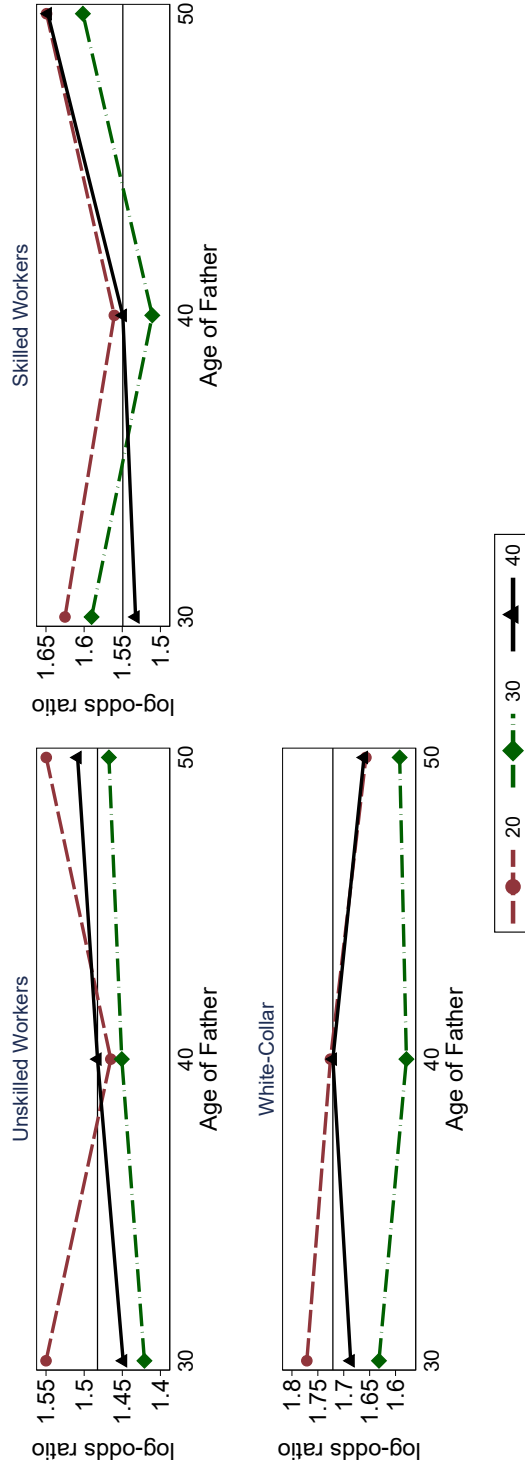FIGURE 20. CORRELATION COEFFICIENT OF THE STANDARDIZED HISCAM MEASURE BY AGE AT CLASSIFICATION.
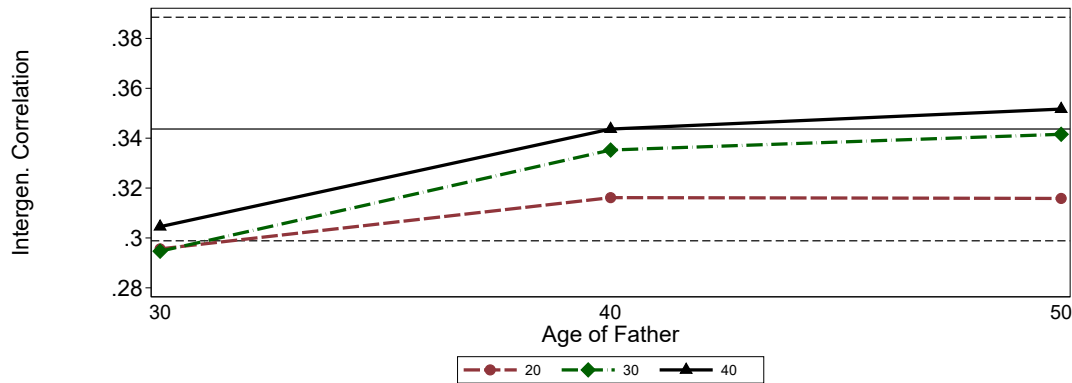
Note: The age at classification of the son is color coded, the father's age at classification is on the x-axis. The solid (dashed) black line represents the estimate (confidence intervals) of the 40-40 sample (both son and father categorized at age 40).

FIGURE 21. SHARE OF (UPWARD AND DOWNWARD) MOBILE ACCORDING TO SEP CATEGORIES BY LINKING PROCEDURE.
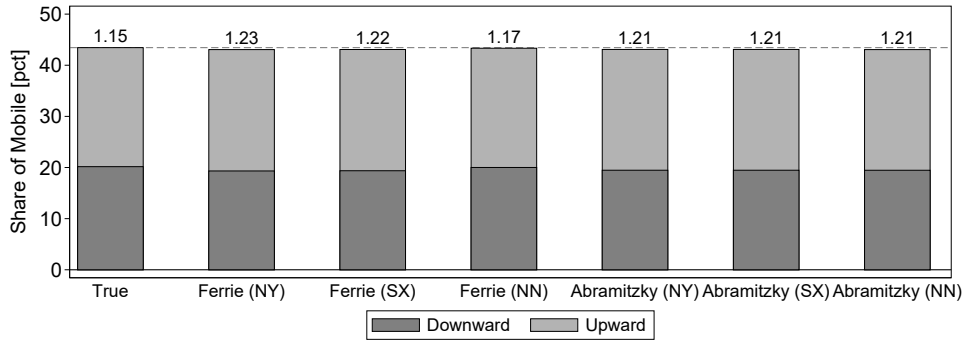
Note: The height of the bars displays the share of mobile sons that did not enter the same occupational category as their father (off-diagonal in the transition matrices). Upward mobile individuals enter higher SEP than their father. The numbers above the bars denote the ratio of upward to downward mobility. The horizontal line marks the share of mobile in the baseline sample ("True").

FIGURE 22. SHARE OF (UPWARD AND DOWNWARD) MOBILE ACCORDING TO LONG-FERRIE CATEGORIES BY LINKING PROCEDURE.

Note: The height of the bars displays the share of mobile sons that did not enter the same occupational category as their father (off-diagonal in the transition matrices). Upward mobility denotes individuals that move closer to white-collar occupations. The numbers above the bars denote the ratio of upward to downward mobility. The horizontal line marks the share of mobile in the baseline sample ("True").

FIGURE 23. SHARE OF (UPWARD AND DOWNWARD) MOBILE ACCORDING TO EXTENDED
LONG-FERRIE CATEGORIES BY LINKING PROCEDURE.

Note: The height of the bars displays the share of mobile sons that did not enter the same occupational category as their father (off-diagonal in the transition matrices). Upward mobility denotes individuals that move closer to higher white-collar occupations. The numbers above the bars denote the ratio of upward to downward mobility. The horizontal line marks the share of mobile in the baseline sample ("True").

FIGURE 24. TWO-WAY LOG-ODDS RATIOS ACCORDING TO SEP CATEGORIES BY LINKING PROCEDURE.
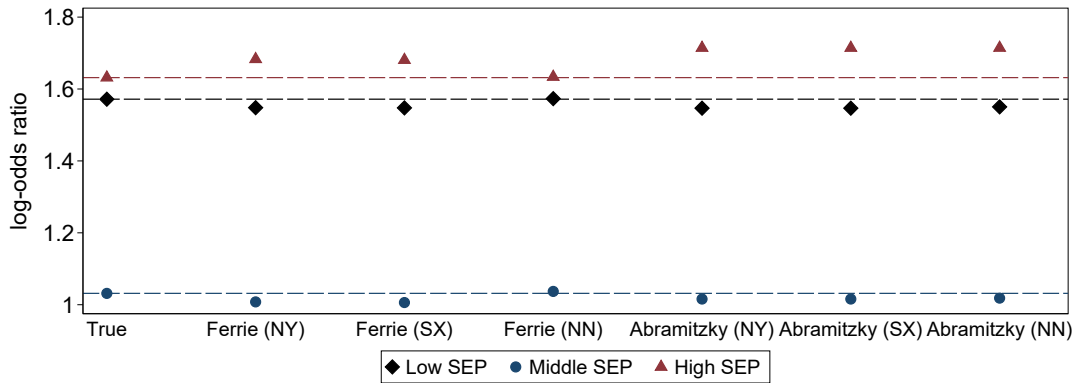
Note: The dashed lines denote the level of the log-odds ratios in the baseline sample (color-coded, "True").

FIGURE 25. TWO-WAY LOG-ODDS RATIOS ACCORDING TO LONG-FERRIE CATEGORIES BY LINKING PROCEDURE.
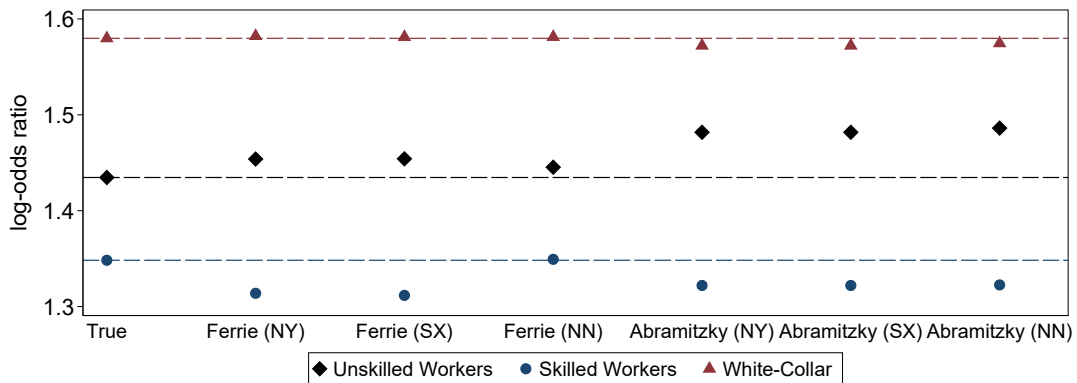
Note: The dashed lines denote the level of the log-odds ratios in the baseline sample (color-coded, "True").

FIGURE 26. TWO-WAY LOG-ODDS RATIOS ACCORDING TO EXTENDED LONG-FERRIE CATEGORIES BY LINKING PROCEDURE.
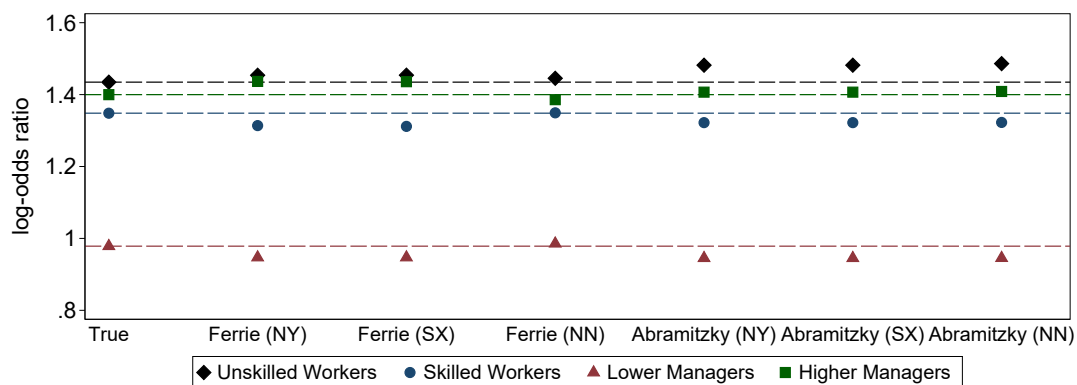
Note: The dashed lines denote the level of the log-odds ratios in the baseline sample (color-coded, "True").

FIGURE 27. ALTHAM STATISTIC ACCORDING TO SEP CATEGORIES BY LINKING PROCEDURE.

Note: This figure contains the Altham statistic controlled for a quadratic function of the father's and the son's age. The lines mark the level (solid) and confidence intervals (dashed) for the baseline sample ("True"). Confidence intervals are calculated by a bootstrapping procedure as explained in Modalsli (2015, p. 8).

Note: This figure contains the Altham statistic controlled for a quadratic function of the father's and the son's age. The lines mark the level (solid) and confidence intervals (dashed) for the baseline sample ("True"). Confidence intervals are calculated by a bootstrapping procedure as explained in Modalsli (2015, p. 8).

FIGURE 29. ALTHAM STATISTIC ACCORDING TO EXTENDED LONG-FERRIE CATEGORIES BY LINKING PROCEDURE.
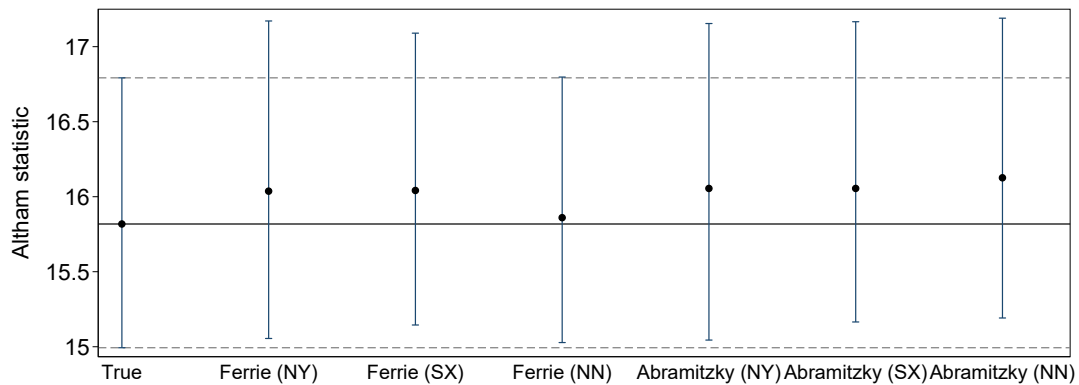
Note: This figure contains the Altham statistic controlled for a quadratic function of the father's and the son's age. The lines mark the level (solid) and confidence intervals (dashed) for the baseline sample ("True"). Confidence intervals are calculated by a bootstrapping procedure as explained in Modalsli (2015, p. 8).

FIGURE 30. CORRELATION COEFFICIENT OF THE STANDARDIZED HISCAM MEASURE BY LINKING PROCEDURE.
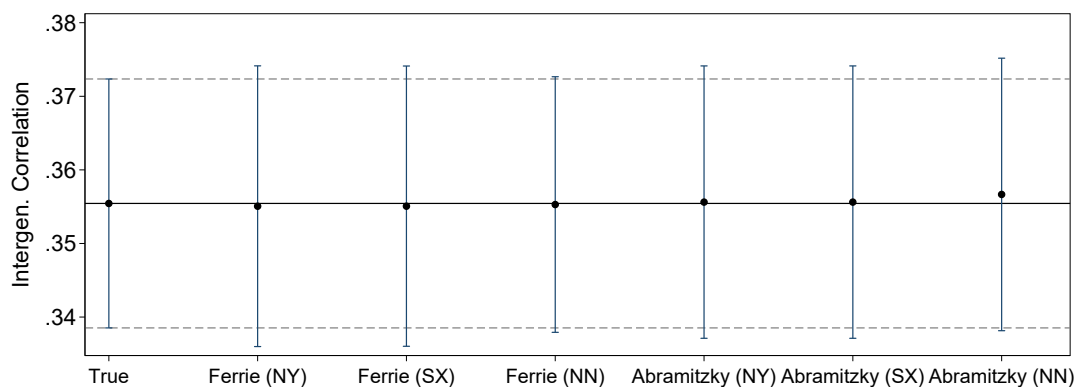
Note: The solid (dashed) black line represents the estimate (confidence intervals) for the baseline sample ("True").

FIGURE 31. RELATIVE SIZE OF BIAS WITH THE STANDARDIZED HISCAM MEASURE.
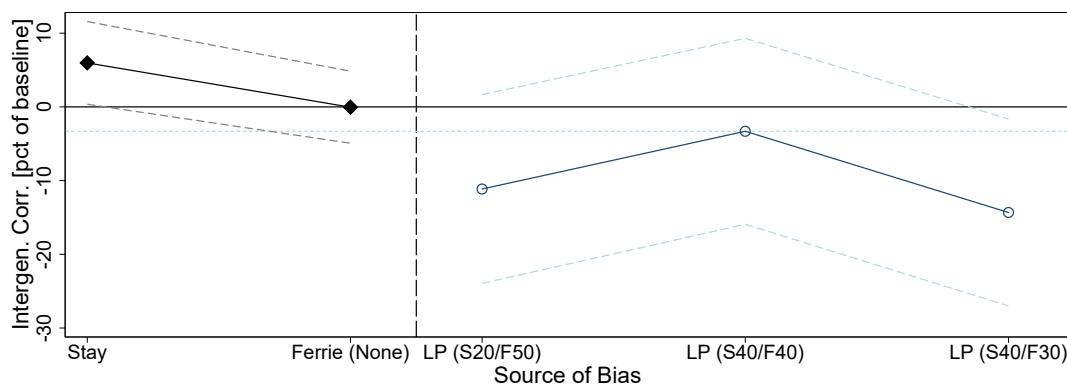
Note: The solid horizontal line represents zero percent bias (baseline sample). The vertical dashed line separates migration and linking procedure as sources of bias on the left-hand side from life pattern differences on the right-hand side. The dashed blue horizontal line marks the level of bias when only including father-son pairs within a five year range around the age of 40. Stay denotes the sample with geographically immobile sons. Ferrie (None) is the linked sample employing Ferrie (1996) without name cleaning. LP is short for life pattern. These samples include father-son pairs that contain information at every age between 20 and 40 (son) and 30 and 50 (father). LP (Sx/Fy) classifies the LP father-son pairs around the age of x (son) and y (father).

Table 1. Descriptives of the baseline sample.

Characteristic	Son	Father
Number of observations	11,384	
Age	35.53	47.17
Low SEP [pct]	37.87	40.72
Middle SEP [pct]	46.56	42.33
High SEP [pct]	15.57	16.95
Unskilled workers [pct]	11.47	14.25
Skilled workers [pct]	24.40	25.39
White-collar [pct]	64.13	60.36
Lower managers [pct]	39.81	33.01
Higher managers [pct]	24.32	27.35

Note: The number of observations refers to the number of father-son pairs. Age is the age at observed occupation closest to forty. The remainder of the table describes the distribution across occupational classes in percent. Lower managers and higher managers are encompassed in the white-collar group.

TABLE 2. DESCRIPTIVES OF FATHER-SON PAIRS BY MIGRATION STATUS.

Characteristic	Stay	Return	Emigrated
Number of observations	8,002	2,126	1,256
Age	35.25	37.21	34.60
Low SEP [pct]	38.54	39.42	30.94
Middle SEP [pct]	44.93	46.84	56.56
High SEP [pct]	16.53	13.75	12.50
Unskilled workers [pct]	12.28	10.12	8.64
Skilled workers [pct]	24.41	26.80	20.23
White-collar [pct]	63.31	63.08	71.13
Lower managers [pct]	38.12	39.42	51.21
Higher managers [pct]	25.18	23.66	19.92
Age (f)	48.47	44.59	44.19
Low SEP [pct] (f)	42.42	37.27	35.31
Middle SEP [pct] (f)	41.81	43.18	44.33
High SEP [pct] (f)	15.77	19.55	20.36
Unskilled workers [pct] (f)	15.52	11.12	11.20
Skilled workers [pct] (f)	25.84	25.66	22.03
White-collar [pct] (f)	58.63	63.22	66.77
Lower managers [pct] (f)	32.23	33.71	36.87
Higher managers [pct] (f)	26.41	29.51	29.90

Note: The number of observations refers to the number of father-son pairs by migration status. Age is the age at observed occupation closest to forty. The remainder of the table describes the distribution across occupational classes in percent. Lower managers and higher managers are encompassed in the white-collar group. Tables 3 and 30 split the migrants by destination. Rows with an (f) capture the values for the fathers, those without an (f) capture the values for the sons.

TABLE 3. DESCRIPTIVES OF FATHER-SON PAIRS BY DESTINATION CONTINENT.

Characteristic	Europe	N. America	SA/Afr/As/Aus
Number of observations	2,088	453	358
Age	36.46	36.44	33.11
Low SEP [pct]	21.23	40.85	26.16
Middle SEP [pct]	61.59	51.49	64.58
High SEP [pct]	17.19	7.66	9.26
Unskilled workers [pct]	5.27	9.81	7.30
Skilled workers [pct]	14.36	29.21	15.45
White-collar [pct]	80.37	60.98	77.25
Lower managers [pct]	51.93	46.70	64.33
Higher managers [pct]	28.44	14.29	12.92
Age (f)	44.51	43.71	44.60
Low SEP [pct] (f)	27.70	38.40	31.80
Middle SEP [pct] (f)	49.63	41.91	50.73
High SEP [pct] (f)	22.68	19.69	17.48
Unskilled workers [pct] (f)	8.55	13.14	10.92
Skilled workers [pct] (f)	18.36	24.71	20.35
White-collar [pct] (f)	73.08	62.16	68.73
Lower managers [pct] (f)	39.53	32.75	42.43
Higher managers [pct] (f)	33.55	29.41	26.30

Note: The number of observations refers to the number of father-son pairs by destination continent. Note that one father-son pair may be included in more than one sub-sample because of multiple migration. Age is the age at observed occupation closest to forty. The remainder of the table describes the distribution across occupational classes in percent. Lower managers and higher managers are encompassed in the white-collar group. Table 30 splits the SA/Afr/As/Aus sample in separate parts (South America, Africa, Asia, and Australia). Rows with an (f) capture the values for the fathers, those without an (f) capture the values for the sons.

TABLE 4. DESCRIPTIVES BY AGE AT CLASSIFICATION.

Characteristic	20	30	40	50
Number of ovservations		1,476		
Age	20.30	30.31	39.95	
Low SEP [pct]	36.86	32.99	28.66	
Middle SEP [pct]	44.38	50.75	53.12	
High SEP [pct]	18.77	16.26	18.22	
Unskilled workers [pct]	7.04	7.11	7.25	
Skilled workers [pct]	28.22	24.07	19.63	
White-collar [pct]	64.74	68.82	73.12	
Lower managers [pct]	42.22	42.93	44.26	
Higher managers [pct]	22.52	25.90	28.85	
Age (f)		30.60	40.29	50.01
Low SEP [pct] (f)		37.20	34.82	32.72
Middle SEP [pct] (f)		41.12	42.07	42.55
High SEP [pct] (f)		21.68	23.10	24.73
Unskilled workers [pct] (f)		10.27	9.22	7.53
Skilled workers [pct] (f)		27.73	25.97	24.77
White-collar [pct] (f)		62.00	64.81	67.70
Lower managers [pct] (f)		33.99	34.34	34.27
Higher managers [pct] (f)		28.01	30.47	33.43

Note: The numbers in the columns refer to the approximate age at which I classify the occupations of individuals. The number of observations refers to the number of father-son pairs. Age is the age at observed occupation closest to 20, 30, 40, or 50. The remainder of the table describes the distribution across occupational classes in percent. Lower managers and higher managers are encompassed in the white-collar group. Rows with an (f) capture the values for the fathers, those without an (f) capture the values for the sons.

TABLE 5. DESCRIPTIVES BY LINKING MECHANISM.

Mechanism	Nobs	Gen	Age	L [pct]	M [pct]	H [pct]
Ferrie (NYSIIS)	9,492	Son	34.55	37.59	47.93	14.48
		Father	47.00	41.58	42.99	15.43
Ferrie (Soundex)	9,473	Son	34.54	37.55	47.94	14.50
		Father	47.00	41.56	43.00	15.44
Ferrie (None)	9,492	Son	34.55	37.59	47.93	14.48
		Father	47.00	41.58	42.99	15.43
Abramitzky (NYSIIS)	8,950	Son	34.21	37.74	48.22	14.05
		Father	47.10	41.85	43.30	14.85
Abramitzky (Soundex)	10,806	Son	35.42	37.87	46.78	15.36
		Father	47.08	40.97	42.38	16.65
Abramitzky (None)	8,979	Son	34.21	37.74	48.24	14.02
		Father	47.07	41.81	43.30	14.89

Note: Mechanism denotes the linking mechanism employed. The number of observations (Nobs) refers to the number of father-son pairs (matches). Gen refers to generation and denotes whether the captured values are for fathers or sons. Age is the age at observed occupation closest to forty. L, M, and H are the fraction of individuals with low, middle, and high SEP occupations, respectively. The distribution across Long-Ferrie categories is displayed in Table 6.

TABLE 6. DESCRIPTIVES BY LINKING MECHANISM—DISTRIBUTION ACROSS LONG-FERRIE CATEGORIES.

Mechanism	Gen	U [pct]	S [pct]	W [pct]	LW [pct]	HW [pct]
Ferrie (NYSIIS)	Son	11.53	23.80	64.66	41.19	23.47
	Father	14.55	25.83	59.62	34.24	25.38
Ferrie (Soundex)	Son	11.54	23.74	64.72	41.23	23.48
	Father	14.55	25.77	59.67	34.27	25.40
Ferrie (None)	Son	11.53	23.80	64.66	41.19	23.47
	Father	14.55	25.83	59.62	34.24	25.38
Abramitzky (NYSIIS)	Son	11.53	23.86	64.61	41.39	23.22
	Father	14.70	25.87	59.44	34.48	24.96
Abramitzky (Soundex)	Son	11.53	24.24	64.24	40.03	24.21
	Father	14.34	25.46	60.20	33.33	26.87
Abramitzky (None)	Son	11.54	23.84	64.62	41.41	23.21
	Father	14.67	25.86	59.48	34.47	25.00

Note: Mechanism denotes the linking mechanism employed. The number of observations and average age is displayed in Table 5. Gen refers to generation and denotes whether the captured values are for fathers or sons. U, S, W, LW, and HW are the fraction of individuals in the occupational group of unskilled workers, skilled workers, white-collar, lower managers, and higher managers.

Table 7. Evaluation of linking procedures.

Procedure	Nobs	Matches	Correct	Type I
Ferrie (NYSIIS)		10,551 (82.49)	10,548 (82.46)	2.84
Ferrie (Soundex)		10,526 (82.29)	10,523 (82.27)	2.85
Ferrie (None)	12,791	10,551 (82.49)	10,548 (82.46)	2.84
Abramitzky (NYSIIS)		9,922 (77.57)	9,920 (77.55)	2.02
Abramitzky (Soundex)		12,082 (94.46)	12,080 (94.44)	1.66
Abramitzky (None)		9,955 (77.83)	9,953 (77.81)	2.01

Note: Procedure denotes the linking mechanism employed. Nobs denotes the number of observations in the baseline sample (without excluding farmers). Matches is the absolute number of matched father-son pairs (share of total observations [pct] in brackets). Correct is the absolute number of correct matches (share of total observations [pct] in brackets corresponding to 1 – type II error rate). Type I is the share of type I errors per mille.

Table 8. Relative size of bias—SEP classification.

Sample	M	IC	AS	$\Theta_{2,1}$	$\Theta_{2,2}$	$\Theta_{2,3}$
Baseline	0.43	0.36	9.44	1.57	1.03	1.63
Stay	-1.37	5.97	10.05	0.59	1.95	10.42
Ferrie (None)	-0.27	-0.04	0.59	0.10	0.53	0.12
LP (S20/F50)	2.76	-11.14	-14.53	11.98	10.20	-22.24
LP (S40/F40)	1.05	-3.31	-3.14	17.39	11.13	-16.57
LP (S40/F30)	4.01	-14.32	-6.12	13.80	0.59	-21.95

Note: Baseline marks the estimates of the mobility measures in the baseline sample. Stay denotes the sample of geographically immobile sons. Ferrie (None) is the linked sample employing Ferrie (1996) without name cleaning. LP is short for life pattern. These samples include father-son pairs that contain information at every age between 20 and 40 (son) and 30 and 50 (father). LP (Sx/Fy) classifies the LP father-son pairs around the age of x (son) and y (father). M is the share of mobile individuals, IC is the intergenerational correlation coefficient based on HISCAM, AS the controlled Altham statistic, and $\Theta_{2,x}$ the two-way log-odds ratio of category $x \in \{1, 2, 3\}$ corresponding to low, middle, and high SEP. The relative size of the deviation is denoted in percent of the baseline.

TABLE 9. RELATIVE SIZE OF BIAS—LONG-FERRIE CLASSIFICATION.

Sample	M	IC	AS	$\Theta_{2,1}$	$\Theta_{2,2}$	$\Theta_{2,3}$
Baseline	0.37	0.36	7.49	1.43	1.35	1.58
Stay	1.59	5.97	3.85	3.46	-0.37	1.36
Ferrie (None)	0.05	-0.04	0.59	0.75	0.08	0.09
LP (S20/F50)	-13.35	-11.14	4.83	8.02	22.32	4.90
LP (S40/F40)	-13.54	-3.31	4.37	3.33	14.91	8.94
LP (S40/F30)	-9.41	-14.32	2.88	0.99	13.67	6.76

Note: Baseline marks the estimates of the mobility measures in the baseline sample. Stay denotes the sample of geographically immobile sons. Ferrie (None) is the linked sample employing Ferrie (1996) without name cleaning. LP is short for life pattern. These samples include father-son pairs that contain information at every age between 20 and 40 (son) and 30 and 50 (father). LP (Sx/Fy) classifies the LP father-son pairs around the age of x (son) and y (father). M is the share of mobile individuals, IC is the intergenerational correlation coefficient based on HISCAM, AS the controlled Altham statistic, and $\Theta_{2,x}$ the two-way log-odds ratio of category $x \in \{1, 2, 3\}$ corresponding to unskilled, skilled, and white-collar workers. The relative size of the deviation is denoted in percent of the baseline.

TABLE 10. RELATIVE SIZE OF BIAS—EXTENDED LONG-FERRIE CLASSIFICATION.

Sample	M	IC	AS	$\Theta_{2,1}$	$\Theta_{2,2}$	$\Theta_{2,3}$	$\Theta_{2,4}$
Baseline	0.54	0.36	15.82	1.43	1.35	0.98	1.40
Stay	-0.56	5.97	5.47	3.46	-0.37	1.06	7.99
Ferrie (None)	-0.05	-0.04	0.27	0.75	0.08	0.72	-1.00
LP (S20/F50)	-4.07	-11.14	-6.67	8.02	22.32	9.65	-13.95
LP (S40/F40)	-2.77	-3.31	-4.11	3.33	14.91	12.58	-20.64
LP (S40/F30)	-0.55	-14.32	-5.57	0.99	13.67	7.96	-27.52

Note: Baseline marks the estimates of the mobility measures in the baseline sample. Stay denotes the sample of geographically immobile sons. Ferrie (None) is the linked sample employing Ferrie (1996) without name cleaning. LP is short for life pattern. These samples include father-son pairs that contain information at every age between 20 and 40 (son) and 30 and 50 (father). LP (Sx/Fy) classifies the LP father-son pairs around the age of x (son) and y (father). M is the share of mobile individuals, IC is the intergenerational correlation coefficient based on HISCAM, AS the controlled Altham statistic, and $\Theta_{2,x}$ the two-way log-odds ratio of category $x \in \{1, 2, 3, 4\}$ corresponding to unskilled, skilled, lower, and higher white-collar workers. The relative size of the deviation is denoted in percent of the baseline.

A Appendix

A.1 Transition Matrices

This section contains all transition matrices referred to in the main section of this paper.

TABLE 11. TRANSITION MATRICES OF SEP—BASELINE SAMPLE.

Son's occupation	Father's occupation			Column sum
	L	M	H	
low SEP (L)	2,764	1,138	405	4,307
middle SEP (M)	1,610	2,946	753	5,309
high SEP (H)	309	731	728	1,768
Row sum	4,683	4,815	1,886	11,384

Note: The baseline sample conveys all father-son pairs available in the data.

TABLE 12. TRANSITION MATRICES OF LONG-FERRIE CATEGORIES—BASELINE SAMPLE.

Son's occupation	Father's occupation			Column sum
	W	S	U	
White-Collar (W)	5,199	1,226	681	7,106
Skilled Workers (S)	957	1,269	477	2,703
Unskilled Workers (U)	464	335	449	1,248
Row sum	6,620	2,830	1,607	11,057

Note: The baseline sample conveys all father-son pairs available in the data.

TABLE 13. TRANSITION MATRICES OF EXTENDED LONG-FERRIE CATEGORIES—BASELINE SAMPLE.

Son's occupation	Father's occupation				Column sum
	H	L	S	U	
Higher Managers (H)	1,341	793	382	182	2,698
Lower Managers (L)	1,037	2,028	844	499	4,408
Skilled Workers (S)	414	543	1,269	477	2,703
Unskilled Workers (U)	187	277	335	449	1,248
Row sum	2,979	3,641	2,830	1,607	11,057

Note: The baseline sample conveys all father-son pairs available in the data.

TABLE 14. TRANSITION MATRICES OF SEP—BY GEOGRAPHIC MOBILITY.

(A). Stayers (geographically immobile).

Son's occupation	Father's occupation			Column sum
	L	M	H	
low SEP (L)	2,030	799	250	3,079
middle SEP (M)	1,161	1,999	441	3,601
high SEP (H)	241	537	544	1,322
Row sum	3,432	3,335	1,235	8,002

(B). Return migrants.

Son's occupation	Father's occupation			Column sum
	L	M	H	
low SEP (L)	502	231	105	838
middle SEP (M)	256	564	180	1,000
high SEP (H)	44	125	119	288
Row sum	802	920	404	2,126

(C.) Emigrants.

Son's occupation	Father's occupation			Column sum
	L	M	H	
low SEP (L)	232	108	50	390
middle SEP (M)	193	383	132	708
high SEP (H)	24	69	65	158
Row sum	449	560	247	1,256

Note: The sample name refers to the emigration status of the son in the father-son pair.

TABLE 15. TRANSITION MATRICES OF SEP—MIGRANTS BY DESTINATION.

(A). Europe.

| Son's occupation | Father's occupation | | | Column |
	L	M	H	sum
low SEP (L)	237	143	68	448
middle SEP (M)	304	739	241	1,284
high SEP (H)	45	159	152	356
Row sum	586	1,041	461	2,088

(B). North America.

| Son's occupation | Father's occupation | | | Column |
	L	M	H	sum
low SEP (L)	107	53	25	185
middle SEP (M)	59	126	48	233
high SEP (H)	8	13	14	35
Row sum	174	192	87	453

(C). South America, Australia, Asia, or Africa.

| Son's occupation | Father's occupation | | | Column |
	L	M	H	sum
low SEP (L)	52	34	8	94
middle SEP (M)	55	140	37	232
high SEP (H)	6	16	10	32
Row sum	113	190	55	358

Note: These tables only contain father-son pairs with geographically mobile sons. The sample name refers to the destination continent(s).

(A). Stayers (geographically immobile).

| Son's occupation | Father's occupation | | | Column |
	W	S	U	sum
White-Collar (W)	3,530	871	509	4,910
Skilled Workers (S)	637	898	358	1,893
Unskilled Workers (U)	325	243	366	934
Row sum	4,492	2,012	1,233	7,737

(B). Return migrants.

| Son's occupation | Father's occupation | | | Column |
	W	S	U	sum
White-Collar (W)	998	219	101	1,318
Skilled Workers (S)	217	265	75	557
Unskilled Workers (U)	92	59	58	209
Row sum	1,307	543	234	2,084

(C). Emigrants.

| Son's occupation | Father's occupation | | | Column |
	W	S	U	sum
White-Collar (W)	671	136	71	878
Skilled Workers (S)	103	106	44	253
Unskilled Workers (U)	47	33	25	105
Row sum	821	275	140	1,236

Note: The sample name refers to the emigration status of the son in the father-son pair.

TABLE 17. TRANSITION MATRICES OF LONG-FERRIE CATEGORIES—MIGRANTS BY DESTINATION.

(A). Europe.

| Son's occupation | Father's occupation | | | Column |
	W	S	U	sum
White-Collar (W)	1,305	228	111	1,644
Skilled Workers (S)	130	128	41	299
Unskilled Workers (U)	53	27	27	107
Row sum	1,488	383	179	2,050

(B). North America.

| Son's occupation | Father's occupation | | | Column |
	W	S	U	sum
White-Collar (W)	204	44	25	273
Skilled Workers (S)	54	55	23	132
Unskilled Workers (U)	24	10	10	44
Row sum	282	109	58	449

(C). South America, Africa, Asia, or Australia.

| Son's occupation | Father's occupation | | | Column |
	W	S	U	sum
White-Collar (W)	204	38	22	264
Skilled Workers (S)	22	20	11	53
Unskilled Workers (U)	13	9	3	25
Row sum	239	67	36	342

Note: These tables only contain father-son pairs with geographically mobile sons. The sample name refers to the destination continent(s).

TABLE 18. TRANSITION MATRICES OF EXTENDED LONG-FERRIE CATEGORIES—BY GEOGRAPHIC MOBILITY.

(A). Stayers (geographically immobile).

Son's occupation	Father's occupation				Column
	H	L	S	U	sum
Higher Managers (H)	976	554	280	147	1,957
Lower Managers (L)	657	1,343	591	362	2,953
Skilled Workers (S)	257	380	898	358	1,893
Unskilled Workers (U)	126	199	243	366	934
Row sum	2,016	2,476	2,012	1,233	7,737

(B). Return migrants.

Son's occupation	Father's occupation				Column
	H	L	S	U	sum
Higher Managers (H)	239	156	75	22	492
Lower Managers (L)	222	381	144	79	826
Skilled Workers (S)	100	117	265	75	557
Unskilled Workers (U)	43	49	59	58	209
Row sum	604	703	543	234	2,084

(C). Emigrants.

Son's occupation	Father's occupation				Column
	H	L	S	U	sum
Higher Managers (H)	126	83	27	13	249
Lower Managers (L)	158	304	109	58	629
Skilled Workers (S)	57	46	106	44	253
Unskilled Workers (U)	18	29	33	25	105
Row sum	359	462	275	140	1,236

Note: The sample name refers to the emigration status of the son in the father-son pair.

TABLE 19. TRANSITION MATRICES OF EXTENDED LONG-FERRIE CATEGORIES—MIGRANTS BY DESTINATION.

(A). Europe.

Son's occupation	Father's occupation				Column sum
	H	L	S	U	
Higher Managers (H)	300	190	72	22	584
Lower Managers (L)	278	537	156	89	1,060
Skilled Workers (S)	67	63	128	41	299
Unskilled Workers (U)	29	24	27	27	107
Row sum	674	814	383	179	2,050

(B). North America.

Son's occupation	Father's occupation				Column sum
	H	L	S	U	
Higher Managers (H)	28	17	9	9	63
Lower Managers (L)	63	96	35	16	210
Skilled Workers (S)	26	28	55	23	132
Unskilled Workers (U)	11	13	10	10	44
Row sum	128	154	109	58	449

(C). South America, Africa, Asia, or Australia.

Son's occupation	Father's occupation				Column sum
	H	L	S	U	
Higher Managers (H)	20	19	4	2	45
Lower Managers (L)	57	108	34	20	219
Skilled Workers (S)	7	15	20	11	53
Unskilled Workers (U)	3	10	9	3	25
Row sum	87	152	67	36	342

Note: These tables only contain father-son pairs with geographically mobile sons. The sample name refers to the destination continent(s).

TABLE 20. TRANSITION MATRICES OF SEP—BY AGE AT CLASSIFICATION.

(A). Son categorized around 20, father categorized around 30.

| Son's occupation | Father's occupation | | | Column |
	L	M	H	sum
low SEP (L)	350	121	73	544
middle SEP (M)	151	379	125	655
high SEP (H)	48	107	122	277
Row sum	549	607	320	1,476

(B). Son categorized around 20, father categorized around 40.

| Son's occupation | Father's occupation | | | Column |
	L	M	H	sum
low SEP (L)	331	130	83	544
middle SEP (M)	139	383	133	655
high SEP (H)	44	108	125	277
Row sum	514	621	341	1,476

(C). Son categorized around 20, father categorized around 50.

| Son's occupation | Father's occupation | | | Column |
	L	M	H	sum
low SEP (L)	310	146	88	544
middle SEP (M)	129	378	148	655
high SEP (H)	44	104	129	277
Row sum	483	628	365	1,476

(D). Son categorized around 30, father categorized around 30.

| Son's occupation | Father's occupation | | | Column |
	L	M	H	sum
low SEP (L)	312	107	68	487
middle SEP (M)	197	411	141	749
high SEP (H)	40	89	111	240
Row sum	549	607	320	1,476

(E). Son categorized around 30, father categorized around 40.

| Son's occupation | Father's occupation | | | Column |
	L	M	H	sum
low SEP (L)	299	110	78	487
middle SEP (M)	179	424	146	749
high SEP (H)	36	87	117	240
Row sum	514	621	341	1,476

(F). Son categorized around 30, father categorized around 50.

Son's occupation	Father's occupation			Column sum
	L	M	H	
low SEP (L)	284	127	76	487
middle SEP (M)	164	418	167	749
high SEP (H)	35	83	122	240
Row sum	483	628	365	1,476

(G). Son categorized around 40, father categorized around 30.

Son's occupation	Father's occupation			Column sum
	L	M	H	
low SEP (L)	283	85	55	423
middle SEP (M)	221	412	151	784
high SEP (H)	45	110	114	269
Row sum	549	607	320	1,476

(H). Son categorized around 40, father categorized around 40.

Son's occupation	Father's occupation			Column sum
	L	M	H	
low SEP (L)	275	87	61	423
middle SEP (M)	199	429	156	784
high SEP (H)	40	105	124	269
Row sum	514	621	341	1,476

(I). Son categorized around 40, father categorized around 50.

Son's occupation	Father's occupation			Column sum
	L	M	H	
low SEP (L)	261	101	61	423
middle SEP (M)	181	430	173	784
high SEP (H)	41	97	131	269
Row sum	483	628	365	1,476

Note: Sons are categorized around the age of 20 (between 16 and 25), 30 (between 25 and 35), or 40 (between 35 and 45). Fathers are categorized around the age of 30, 40, and 50 +/-5.

TABLE 21. TRANSITION MATRICES OF LONG-FERRIE CATEGORIES—BY AGE AT CLASSIFICATION.

(A). Son categorized around 20, father categorized around 30.

Son's occupation	Father's occupation			Column sum
	W	S	U	
White-Collar (W)	703	158	59	920
Skilled Workers (S)	134	211	56	401
Unskilled Workers (U)	44	25	31	100
Row sum	881	394	146	1,421

(B). Son categorized around 20, father categorized around 40.

Son's occupation	Father's occupation			Column sum
	W	S	U	
White-Collar (W)	723	146	51	920
Skilled Workers (S)	151	197	53	401
Unskilled Workers (U)	47	26	27	100
Row sum	921	369	131	1,421

(C). Son categorized around 20, father categorized around 50.

Son's occupation	Father's occupation			Column sum
	W	S	U	
White-Collar (W)	741	132	47	920
Skilled Workers (S)	170	195	36	401
Unskilled Workers (U)	51	25	24	100
Row sum	962	352	107	1,421

(D). Son categorized around 30, father categorized around 30.

Son's occupation	Father's occupation			Column sum
	W	S	U	
White-Collar (W)	723	182	73	978
Skilled Workers (S)	113	185	44	342
Unskilled Workers (U)	45	27	29	101
Row sum	881	394	146	1,421

(E). Son categorized around 30, father categorized around 40.

Son's occupation	Father's occupation			Column sum
	W	S	U	
White-Collar (W)	745	171	62	978
Skilled Workers (S)	128	172	42	342
Unskilled Workers (U)	48	26	27	101
Row sum	921	369	131	1,421

(F). Son categorized around 30, father categorized around 50.

Son's occupation	Father's occupation			Column sum
	W	S	U	
White-Collar (W)	771	154	53	978
Skilled Workers (S)	140	171	31	342
Unskilled Workers (U)	51	27	23	101
Row sum	962	352	107	1,421

(G). Son categorized around 40, father categorized around 30.

Son's occupation	Father's occupation			Column sum
	W	S	U	
White-Collar (W)	755	208	76	1,039
Skilled Workers (S)	85	154	40	279
Unskilled Workers (U)	41	32	30	103
Row sum	881	394	146	1,421

(H). Son categorized around 40, father categorized around 40.

Son's occupation	Father's occupation			Column sum
	W	S	U	
White-Collar (W)	785	190	64	1,039
Skilled Workers (S)	92	148	39	279
Unskilled Workers (U)	44	31	28	103
Row sum	921	369	131	1,421

(I). Son categorized around 40, father categorized around 50.

Son's occupation	Father's occupation			Column sum
	W	S	U	
White-Collar (W)	809	174	56	1,039
Skilled Workers (S)	104	148	27	279
Unskilled Workers (U)	49	30	24	103
Row sum	962	352	107	1,421

Note: Sons are categorized around the age of 20 (between 16 and 25), 30 (between 25 and 35), or 40 (between 35 and 45). Fathers are categorized around the age of 30, 40, and 50 +/-5.

TABLE 22. TRANSITION MATRICES OF EXTENDED LONG-FERRIE CATEGORIES—BY AGE AT CLASSIFICATION.

(A). Son categorized around 20, father categorized around 30.

Son's occupation	Father's occupation				Column sum
	H	L	S	U	
Higher Managers (H)	147	104	49	20	320
Lower Managers (L)	155	297	109	39	600
Skilled Workers (S)	72	62	211	56	401
Unskilled Workers (U)	24	20	25	31	100
Row sum	398	483	394	146	1,421

(B). Son categorized around 20, father categorized around 40.

Son's occupation	Father's occupation				Column sum
	H	L	S	U	
Higher Managers (H)	157	102	41	20	320
Lower Managers (L)	171	293	105	31	600
Skilled Workers (S)	81	70	197	53	401
Unskilled Workers (U)	24	23	26	27	100
Row sum	433	488	369	131	1,421

(C). Son categorized around 20, father categorized around 50.

Son's occupation	Father's occupation				Column sum
	H	L	S	U	
Higher Managers (H)	177	89	36	18	320
Lower Managers (L)	186	289	96	29	600
Skilled Workers (S)	84	86	195	36	401
Unskilled Workers (U)	28	23	25	24	100
Row sum	475	487	352	107	1,421

(D). Son categorized around 30, father categorized around 30.

Son's occupation	Father's occupation				Column sum
	H	L	S	U	
Higher Managers (H)	164	115	65	24	368
Lower Managers (L)	149	295	117	49	610
Skilled Workers (S)	61	52	185	44	342
Unskilled Workers (U)	24	21	27	29	101
Row sum	398	483	394	146	1,421

(E). Son categorized around 30, father categorized around 40.

Son's occupation	Father's occupation				Column sum
	H	L	S	U	
Higher Managers (H)	177	110	60	21	368
Lower Managers (L)	161	297	111	41	610
Skilled Workers (S)	69	59	172	42	342
Unskilled Workers (U)	26	22	26	27	101
Row sum	433	488	369	131	1,421

(F). Son categorized around 30, father categorized around 50.

Son's occupation	Father's occupation				Column sum
	H	L	S	U	
Higher Managers (H)	200	96	54	18	368
Lower Managers (L)	177	298	100	35	610
Skilled Workers (S)	67	73	171	31	342
Unskilled Workers (U)	31	20	27	23	101
Row sum	475	487	352	107	1,421

(G). Son categorized around 40, father categorized around 30.

Son's occupation	Father's occupation				Column sum
	H	L	S	U	
Higher Managers (H)	178	129	77	26	410
Lower Managers (L)	152	296	131	50	629
Skilled Workers (S)	46	39	154	40	279
Unskilled Workers (U)	22	19	32	30	103
Row sum	398	483	394	146	1,421

(H). Son categorized around 40, father categorized around 40.

Son's occupation	Father's occupation				Column sum
	H	L	S	U	
Higher Managers (H)	197	124	65	24	410
Lower Managers (L)	162	302	125	40	629
Skilled Workers (S)	51	41	148	39	279
Unskilled Workers (U)	23	21	31	28	103
Row sum	433	488	369	131	1,421

(I). Son categorized around 40, father categorized around 50.

Son's occupation	Father's occupation				Column sum
	H	L	S	U	
Higher Managers (H)	220	110	61	19	410
Lower Managers (L)	178	301	113	37	629
Skilled Workers (S)	51	53	148	27	279
Unskilled Workers (U)	26	23	30	24	103
Row sum	475	487	352	107	1,421

Note: Sons are categorized around the age of 20 (between 16 and 25), 30 (between 25 and 35), or 40 (between 35 and 45). Fathers are categorized around the age of 30, 40, and 50 +/-5.

(A). Ferrie (NYSIIS).

Son's occupation	Father's occupation			Column
	L	M	H	sum
low SEP (L)	2,312	974	283	3,569
middle SEP (M)	1,422	2,541	591	4,554
high SEP (H)	245	575	549	1,369
Row sum	3,979	4,090	1,423	9,492

(B). Ferrie (Soundex).

Son's occupation	Father's occupation			Column
	L	M	H	sum
low SEP (L)	2,307	971	282	3,560
middle SEP (M)	1,416	2,538	591	4,545
high SEP (H)	245	574	549	1,368
Row sum	3,968	4,083	1,422	9,473

(C). Ferrie (None).

Son's occupation	Father's occupation			Column
	L	M	H	sum
low SEP (L)	2,312	974	283	3,569
middle SEP (M)	1,422	2,541	591	4,554
high SEP (H)	245	575	549	1,369
Row sum	3,979	4,090	1,423	9,492

Note: The sample name refers to the linking procedure employed to match fathers and sons.

TABLE 25. TRANSITION MATRICES OF SEP—BY NAME CLEANING PROCEDURE (ABRAMITZKY).

(A). Abramitzky (NYSIIS).

| Son's occupation | Father's occupation | | | Column |
	L	M	H	sum
low SEP (L)	2,196	916	266	3,378
middle SEP (M)	1,360	2,414	548	4,322
high SEP (H)	218	546	486	1,250
Row sum	3,774	3,876	1,300	8,950

(B). Abramitzky (Soundex).

| Son's occupation | Father's occupation | | | Column |
	L	M	H	sum
low SEP (L)	2,638	1,080	376	4,094
middle SEP (M)	1,540	2,813	706	5,059
high SEP (H)	293	687	673	1,653
Row sum	4,471	4,580	1,755	10,806

(C). Abramitzky (None).

| Son's occupation | Father's occupation | | | Column |
	L	M	H	sum
low SEP (L)	2,202	920	267	3,389
middle SEP (M)	1,364	2,422	552	4,338
high SEP (H)	218	547	487	1,252
Row sum	3,784	3,889	1,306	8,979

Note: The sample name refers to the linking procedure employed to match fathers and sons.

TABLE 26. TRANSITION MATRICES OF LONG-FERRIE CATEGORIES—BY NAME CLEANING PROCEDURE (FERRIE).

(A). Ferrie (NYSIIS).

| Son's occupation | Father's occupation | | | Column |
	W	S	U	sum
White-Collar (W)	4,299	1,074	576	5,949
Skilled Workers (S)	761	1,039	393	2,193
Unskilled Workers (U)	376	280	390	1,046
Row sum	5,436	2,393	1,359	9,188

(B). Ferrie (Soundex).

| Son's occupation | Father's occupation | | | Column |
	W	S	U	sum
White-Collar (W)	4,297	1,071	573	5,941
Skilled Workers (S)	758	1,034	393	2,185
Unskilled Workers (U)	376	278	390	1,044
Row sum	5,431	2,383	1,356	9,170

(C). Ferrie (None).

| Son's occupation | Father's occupation | | | Column |
	W	S	U	sum
White-Collar (W)	4,299	1,074	576	5,949
Skilled Workers (S)	761	1,039	393	2,193
Unskilled Workers (U)	376	280	390	1,046
Row sum	5,436	2,393	1,359	9,188

Note: The sample name refers to the linking procedure employed to match fathers and sons.

TABLE 27. TRANSITION MATRICES OF LONG-FERRIE CATEGORIES—BY NAME CLEANING PROCEDURE (ABRAMITZKY).

(A). Abramitzky (NYSIIS).

| | Father's occupation | | | Column |
Son's occupation	W	S	U	sum
White-Collar (W)	4,042	1,008	548	5,598
Skilled Workers (S)	714	979	377	2,070
Unskilled Workers (U)	349	271	364	984
Row sum	5,105	2,258	1,289	8,652

(B). Abramitzky (Soundex).

| | Father's occupation | | | Column |
Son's occupation	W	S	U	sum
White-Collar (W)	4,923	1,173	646	6,742
Skilled Workers (S)	893	1,199	453	2,545
Unskilled Workers (U)	445	317	434	1,196
Row sum	6,261	2,689	1,533	10,483

(C). Abramitzky (None).

| | Father's occupation | | | Column |
Son's occupation	W	S	U	sum
White-Collar (W)	4,056	1,012	549	5,617
Skilled Workers (S)	717	981	377	2,075
Unskilled Workers (U)	350	273	365	988
Row sum	5,123	2,266	1,291	8,680

Note: The sample name refers to the linking procedure employed to match fathers and sons.

(A). Ferrie (NYSIIS).

| Son's occupation | Father's occupation | | | | Column |
	H	L	S	U	sum
Higher Managers (H)	1,025	662	328	144	2,159
Lower Managers (L)	844	1,768	746	432	3,790
Skilled Workers (S)	296	465	1,039	393	2,193
Unskilled Workers (U)	134	242	280	390	1,046
Row sum	2,299	3,137	2,393	1,359	9,188

(B). Ferrie (Soundex).

| Son's occupation | Father's occupation | | | | Column |
	H	L	S	U	sum
Higher Managers (H)	1,025	661	328	142	2,156
Lower Managers (L)	844	1,767	743	431	3,785
Skilled Workers (S)	294	464	1,034	393	2,185
Unskilled Workers (U)	134	242	278	390	1,044
Row sum	2,297	3,134	2,383	1,356	9,170

(C). Ferrie (None).

| Son's occupation | Father's occupation | | | | Column |
	H	L	S	U	sum
Higher Managers (H)	1,025	662	328	144	2,159
Lower Managers (L)	844	1,768	746	432	3,790
Skilled Workers (S)	296	465	1,039	393	2,193
Unskilled Workers (U)	134	242	280	390	1,046
Row sum	2,299	3,137	2,393	1,359	9,188

Note: The sample name refers to the linking procedure employed to match fathers and sons.

TABLE 29. TRANSITION MATRICES OF EXTENDED LONG-FERRIE CATEGORIES—BY NAME CLEANING PROCEDURE (ABRAMITZKY).

(A). Abramitzky (NYSIIS).

| Son's occupation | Father's occupation | | | | Column |
	H	L	S	U	sum
Higher Managers (H)	956	622	297	136	2,011
Lower Managers (L)	783	1,681	711	412	3,587
Skilled Workers (S)	269	445	979	377	2,070
Unskilled Workers (U)	125	224	271	364	984
Row sum	2,133	2,972	2,258	1,289	8,652

(B). Abramitzky (Soundex).

| Son's occupation | Father's occupation | | | | Column |
	H	L	S	U	sum
Higher Managers (H)	1,243	758	369	175	2,545
Lower Managers (L)	972	1,950	804	471	4,197
Skilled Workers (S)	385	508	1,199	453	2,545
Unskilled Workers (U)	173	272	317	434	1,196
Row sum	2,773	3,488	2,689	1,533	10,483

(C). Abramitzky (None).

| Son's occupation | Father's occupation | | | | Column |
	H	L	S	U	sum
Higher Managers (H)	959	623	298	136	2,016
Lower Managers (L)	787	1,687	714	413	3,601
Skilled Workers (S)	271	446	981	377	2,075
Unskilled Workers (U)	125	225	273	365	988
Row sum	2,142	2,981	2,266	1,291	8,680

Note: The sample name refers to the linking procedure employed to match fathers and sons.
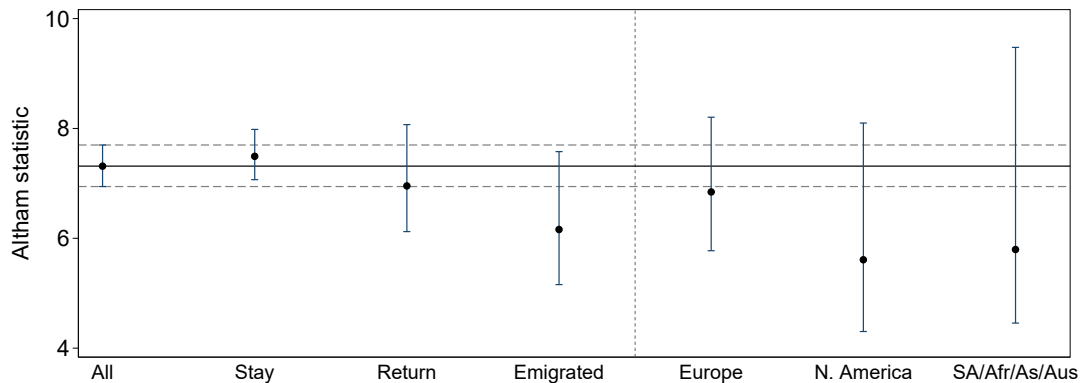
A.2 Supplementary Figures and Tables

FIGURE 32. RAW ALTHAM STATISTIC ACCORDING TO SEP CATEGORIES BY MIGRATION STATUS.
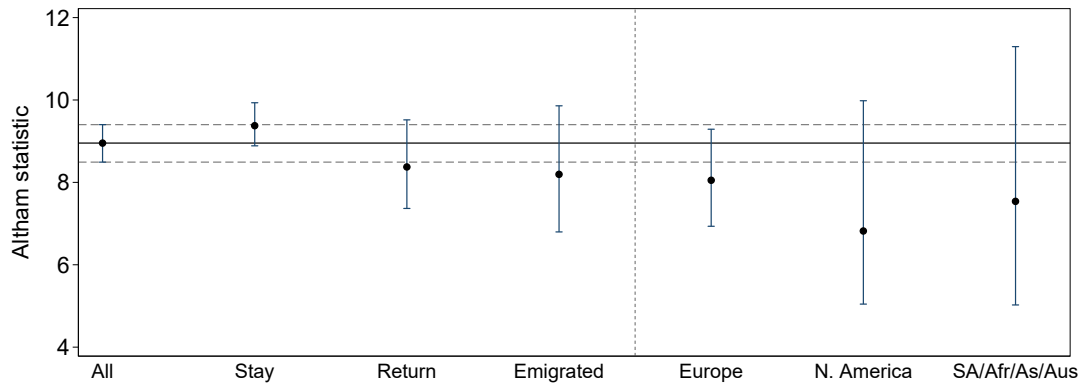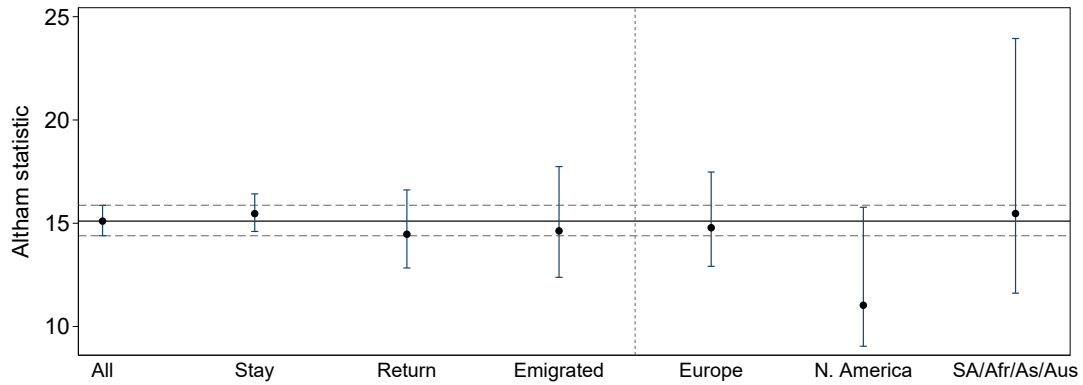

Note: This figure contains the uncontrolled Altham statistic. The lines mark the level (solid) and confidence intervals (dashed) for the baseline sample. Confidence intervals are calculated by a bootstrapping procedure as explained in Modalsli (2015, p. 8). The groups to on the RHS of the vertical dashed line are geographically mobile (Return and Emigrated) split by destination.

FIGURE 33. RAW ALTHAM STATISTIC ACCORDING TO LONG-FERRIE CATEGORIES BY MIGRATION STATUS.
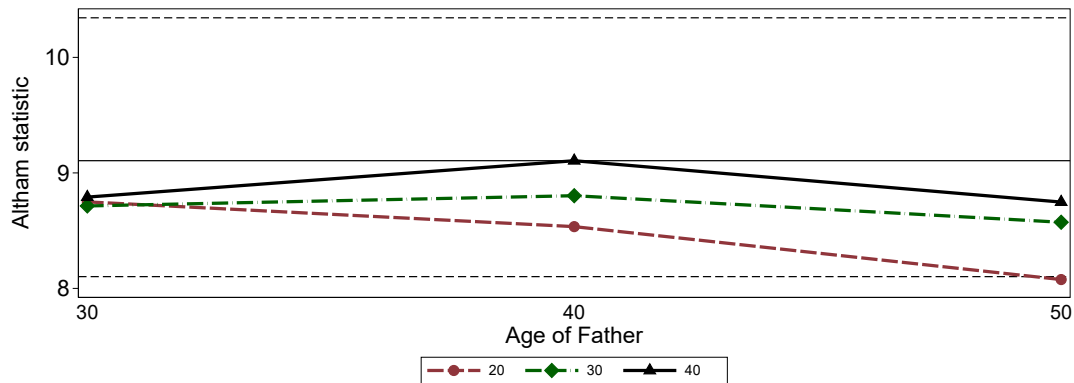

Note: This figure contains the uncontrolled Altham statistic. The lines mark the level (solid) and confidence intervals (dashed) for the baseline sample. Confidence intervals are calculated by a bootstrapping procedure as explained in Modalsli (2015, p. 8). The groups to on the RHS of the vertical dashed line are geographically mobile (Return and Emigrated) split by destination.

Figure 34. Raw Altham statistic according to extended Long-Ferrie categories by migration status.
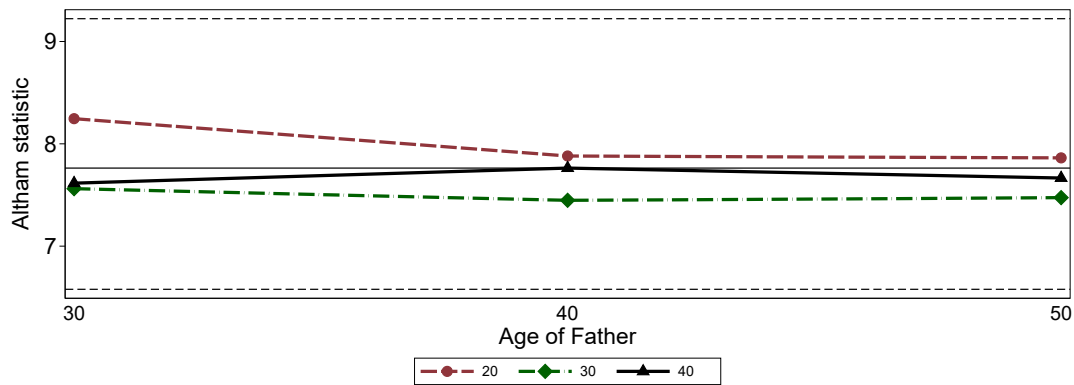
Note: This figure contains the uncontrolled Altham statistic. The lines mark the level (solid) and confidence intervals (dashed) for the baseline sample. Confidence intervals are calculated by a bootstrapping procedure as explained in Modalsli (2015, p. 8). The groups to on the RHS of the vertical dashed line are geographically mobile (Return and Emigrated) split by destination.

Figure 35. Raw Altham statistic according to SEP categories by age at classification.

Note: The age at classification of the son is color coded, the father's age at classification is on the x-axis. This figure contains the uncontrolled Altham statistic. The lines mark the level (solid) and confidence intervals (dashed) of the 40-40 sample (both son and father categorized at age 40). Confidence intervals are calculated by a bootstrapping procedure as explained in Modalsli (2015, p. 8).

FIGURE 36. RAW ALTHAM STATISTIC ACCORDING TO LONG-FERRIE CATEGORIES BY AGE AT CLASSIFICATION.
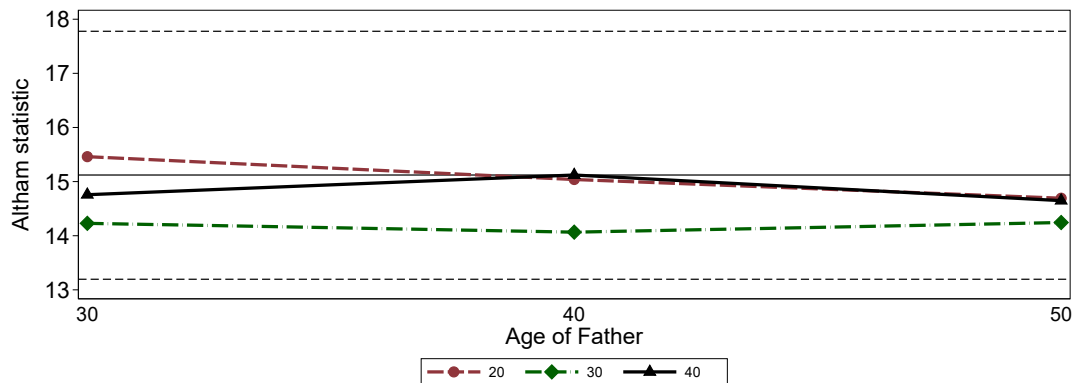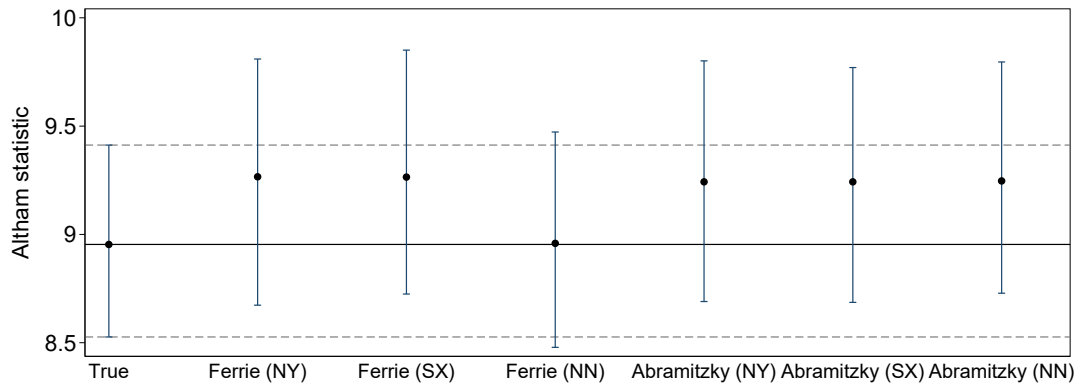
Note: The age at classification of the son is color coded, the father's age at classification is on the x-axis. This figure contains the uncontrolled Altham statistic. The lines mark the level (solid) and confidence intervals (dashed) of the 40-40 sample (both son and father categorized at age 40). Confidence intervals are calculated by a bootstrapping procedure as explained in Modalsli (2015, p. 8).

FIGURE 37. RAW ALTHAM STATISTIC ACCORDING TO EXTENDED LONG-FERRIE CATEGORIES BY AGE AT CLASSIFICATION.

*Note*: The age at classification of the son is color coded, the father's age at classification is on the x-axis. This figure contains the uncontrolled Altham statistic. The lines mark the level (solid) and confidence intervals (dashed) of the 40-40 sample (both son and father categorized at age 40). Confidence intervals are calculated by a bootstrapping procedure as explained in Modalsli (2015, p. 8).

Note: This figure contains the uncontrolled Altham statistic. The lines mark the level (solid) and confidence intervals (dashed) for the baseline sample. Confidence intervals are calculated by a bootstrapping procedure as explained in Modalsli (2015, p. 8).

FIGURE 39. RAW ALTHAM STATISTIC ACCORDING TO LONG-FERRIE CATEGORIES BY LINKING PROCEDURE.
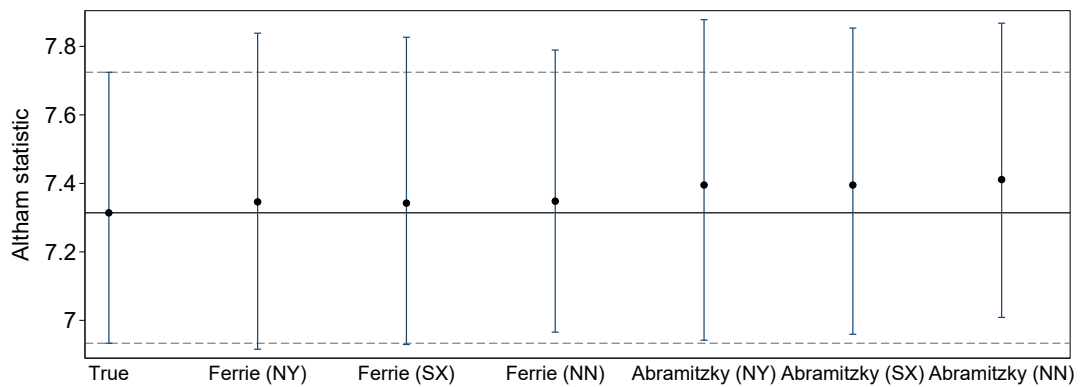
Note: This figure contains the uncontrolled Altham statistic. The lines mark the level (solid) and confidence intervals (dashed) for the baseline sample. Confidence intervals are calculated by a bootstrapping procedure as explained in Modalsli (2015, p. 8).

FIGURE 40. RAW ALTHAM STATISTIC ACCORDING TO EXTENDED LONG-FERRIE CATEGORIES BY LINKING PROCEDURE.
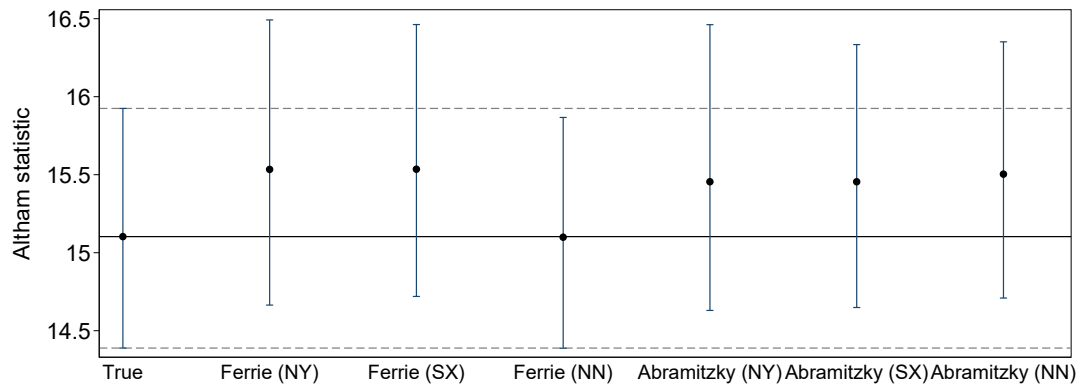
Note: This figure contains the uncontrolled Altham statistic. The lines mark the level (solid) and confidence intervals (dashed) for the baseline sample. Confidence intervals are calculated by a bootstrapping procedure as explained in Modalsli (2015, p. 8).

Characteristic	S. America	Australia	Asia	Africa
Number of observations	146	50	106	21
Age	35.68	34.27	35.40	34.22
Low SEP [pct]	31.33	19.23	10.28	13.04
Middle SEP [pct]	62.00	69.23	80.37	69.57
High SEP [pct]	6.67	11.54	9.35	17.39
Unskilled workers [pct]	5.41	9.80	1.92	4.35
Skilled workers [pct]	21.62	7.84	7.69	8.70
White-collar [pct]	72.97	82.35	90.38	86.96
Lower managers [pct]	62.84	64.71	76.92	73.91
Higher managers [pct]	10.14	17.65	13.46	13.04
Age (f)	44.35	45.29	43.87	49.26
Low SEP [pct] (f)	35.67	26.92	34.78	33.33
Middle SEP [pct] (f)	49.68	50.00	47.83	52.38
High SEP [pct] (f)	14.65	23.08	17.39	14.29
Unskilled workers [pct] (f)	11.46	9.62	13.16	9.52
Skilled workers [pct] (f)	24.84	15.38	20.18	14.29
White-collar [pct] (f)	63.69	75.00	66.67	76.19
Lower managers [pct] (f)	39.49	46.15	38.60	42.86
Higher managers [pct] (f)	24.20	28.85	28.07	33.33

Note: The number of observations refers to the number of father-son pairs by destination continent. Note that one father-son pair may be included in more than one sub-sample because of multiple migration. Age is the age at observed occupation closest to forty. The remainder of the table describes the distribution across occupational classes in percent. Lower managers and higher managers are encompassed in the white-collar group. Rows with an (f) capture the values for the fathers, those without an (f) capture the values for the sons.

(A). South America.

| Son's occupation | Father's occupation | | | Column |
	L	M	H	sum
low SEP (L)	28	13	4	45
middle SEP (M)	22	60	10	92
high SEP (H)	3	2	4	9
Row sum	53	75	18	146

(B). Africa.

| Son's occupation | Father's occupation | | | Column |
	L	M	H	sum
low SEP (L)	2	1	0	3
middle SEP (M)	5	8	2	15
high SEP (H)	0	2	1	3
Row sum	7	11	3	21

(C). Asia.

| Son's occupation | Father's occupation | | | Column |
	L	M	H	sum
low SEP (L)	6	2	3	11
middle SEP (M)	30	46	9	85
high SEP (H)	1	5	4	10
Row sum	37	53	16	106

(D). Australia.

| Son's occupation | Father's occupation | | | Column |
	L	M	H	sum
low SEP (L)	6	3	1	10
middle SEP (M)	8	19	7	34
high SEP (H)	0	3	3	6
Row sum	14	25	11	50

Note: These tables only contain father-son pairs with geographically mobile sons. The sample name refers to the destination continent(s).

TABLE 32. TRANSITION MATRICES OF LONG-FERRIE CATEGORIES—MIGRANTS TO SOUTH AMERICA, AFRICA, ASIA, AND AUSTRALIA.

(A). South America.

| Son's occupation | Father's occupation | | | Column |
	W	S	U	sum
White-Collar (W)	78	18	9	105
Skilled Workers (S)	10	15	6	31
Unskilled Workers (U)	3	4	1	8
Row sum	91	37	16	144

(B). Africa.

| Son's occupation | Father's occupation | | | Column |
	W	S	U	sum
White-Collar (W)	15	1	2	18
Skilled Workers (S)	1	1	0	2
Unskilled Workers (U)	0	1	0	1
Row sum	16	3	2	21

(C). Asia.

| Son's occupation | Father's occupation | | | Column |
	W	S	U	sum
White-Collar (W)	64	18	10	92
Skilled Workers (S)	4	3	1	8
Unskilled Workers (U)	1	0	1	2
Row sum	69	21	12	102

(D). Australia.

| Son's occupation | Father's occupation | | | Column |
	W	S	U	sum
White-Collar (W)	33	5	2	40
Skilled Workers (S)	2	1	1	4
Unskilled Workers (U)	2	1	2	5
Row sum	37	7	5	49

Note: These tables only contain father-son pairs with geographically mobile sons. The sample name refers to the destination continent(s).

TABLE 33. TRANSITION MATRICES OF EXTENDED LONG-FERRIE CATEGORIES—MIGRANTS TO SOUTH AMERICA, AFRICA, ASIA, AND AUSTRALIA.

(A). South America.

Son's occupation	Father's occupation				Column sum
	H	L	S	U	
Higher Managers (H)	7	6	1	0	14
Lower Managers (L)	20	45	17	9	91
Skilled Workers (S)	4	6	15	6	31
Unskilled Workers (U)	1	2	4	1	8
Row sum	32	59	37	16	144

(B). Africa.

Son's occupation	Father's occupation				Column sum
	H	L	S	U	
Higher Managers (H)	1	2	0	0	3
Lower Managers (L)	6	6	1	2	15
Skilled Workers (S)	0	1	1	0	2
Unskilled Workers (U)	0	0	1	0	1
Row sum	7	9	3	2	21

(C). Asia.

Son's occupation	Father's occupation				Column sum
	H	L	S	U	
Higher Managers (H)	9	4	1	0	14
Lower Managers (L)	16	35	17	10	78
Skilled Workers (S)	2	2	3	1	8
Unskilled Workers (U)	0	1	0	1	2
Row sum	27	42	21	12	102

(D). Australia.

Son's occupation	Father's occupation				Column sum
	H	L	S	U	
Higher Managers (H)	5	4	0	0	9
Lower Managers (L)	8	16	5	2	31
Skilled Workers (S)	1	1	1	1	4
Unskilled Workers (U)	0	2	1	2	5
Row sum	14	23	7	5	49

Note: These tables only contain father-son pairs with geographically mobile sons. The sample name refers to the destination continent(s).