Multigenerational persistence: Evidence from 146 years of administrative data

Jørgen Modalsli

February 28, 2021

Abstract

This paper documents multigenerational persistence in economic status, showing that not only do parents influence children's economic outcomes, but so too do grandparents and great-grandparents. Economic persistence is measured using direct grandfather-father-son links, including up to five generations, in administrative data from Norway spanning nearly 150 years (1865 to 2011). The findings are robust to alternative ways of measuring the characteristics of the parent generation, and to alternative indicators of economic status. High persistence is observed also in subsamples where grandchildren had less chance to interact directly with grandparents, suggesting an important role of unexpressed family characteristics in intergenerational transmission. The results indicate a slower occupational convergence across families over time than what is implied by parent-child associations.

Keywords: Multigenerational mobility; human capital transmission; occupational mobility; income mobility; grandfathers

JEL codes: J62, D31, N33, N34

Author affiliation and acknowledgement: Jørgen Modalsli is Head of Department at Oslo Business School, Oslo Metropolitan University and Senior Researcher at Statistics Norway. Email: jorgenmo@oslomet.no. The author would like to thank Rolf Aaberge, Lars Kirkebøen, Andreas Kotsadam, Mikael Lindahl, Kjetil Telle, the Editor, anonymous referees as well as participants at workshops and conferences for helpful comments and discussions.

Disclosure statement: Support from the Norwegian Research Council is acknowledged.

Data replication statement: The data files of individual records from the census files, tax registry and population registry collected less than 100 years ago can only be obtained by authorized researchers according to the guidelines described in http://www.ssb.no/en/omssb/tjenester-og-verktoy/data-til-forskning . There are legal limitations (the Norwegian Statistics Act) on making these data publicly available. However, summary tables of occupational associations are presented in the Online Appendix and programs and do-files used in writing the paper can be made available on request. The use of individual administrative records in this project has been approved by Statistics Norway's Data Protection Officer (personvernombudet) in accordance with ethical and legal regulations.

1 Introduction

How persistent are economic outcomes across generations? To what extent are children's outcomes affected not only by their parents, but also by their grandparents or great-grandparents? An understanding of such persistence is important for understanding the extent of economic mobility over time and can provide information about the relative roles of direct parental involvement and more abstract family human capital in shaping individuals' economic opportunities. This paper demonstrates persistence in economic outcomes over a period of 146 years using a novel linked data set based on six Norwegian full-count censuses between 1865 and 2011 including 167,411 lineages with occupational data on grandfathers, fathers and sons. This represents the largest data set on such multigenerational processes to date.

The results consistently find that grandfathers' economic outcomes predict their grandsons' outcomes. Even after controlling for father's occupation, there are sizeable, statistically significant associations between grandfather and grandson across all time periods studied. The findings are robust to alternative measures including occupational status and income rank. Moreover, measuring the social status of the parent generation in greater detail does not remove the association between grandfather and grandson, ruling out measurement error as the sole factor.

Further, rates of persistence have changed over time in Norway. Persistence is highest in white-collar occupations and among farmers. For white-collar occupations, persistence has decreased over time; the occupational association between a white-collar grandfather and his grandson in 1910 is comparable to that of a white-collar father-son pair in 2011. Despite the decline in white-collar grandfathers' influence over time, it is always the case that having a white-collar grandfather increases the likelihood that one has a white-collar job. Multigenerational persistence among manual skilled workers has also decreased over time, while persistence has increased for farmers and unskilled workers.

The results demonstrate that the existence of multigenerational persistence does not depend on any specific set of economic institutions, as Norwegian society changed dramatically over the time period studied. In 1865, a majority of the population made their living from farming-related activity, and GDP per capita is estimated to have been only around half that of leading European countries (Bolt & van Zanden, 2013). There was no state income tax, and for most of the population only basic elementary education was available. At the end of the period, by contrast, there is a comprehensive welfare system, education at all levels is free, and less than one percent of the population is engaged in farming. Yet across all time periods, grandfathers' occupations have a strong predictive influence on grandsons' occupations.

An exploratory analysis suggests that the influence of grandfathers is not primarily driven by interpersonal interactions. There is strong multi-generational persistence even when grandfather and grandson did not reside in the same region, or when grandfather died at an earlier age. This runs counter to previous results from China (Zeng & Xie, 2014) that attribute multigenerational persistence to direct grandparent-grandchild contact. Hence, the results found in this paper can be reasonably attributed to some characteristics of the family that are not manifested in observable economic outcomes.

This paper, with unique administrative data spanning nearly 150 years, complements and extends the existing literature on long-run economic persistence. Existing work is generally limited by years of data availability, small sample sizes, potential recall error, and other measurement challenges. For example, while Long & Ferrie (2013a) document changes in intergenerational mobility (father-son associations) over time, less is known about whether persistence across multiple generations has changed in a similar manner. Few studies have so far covered a long time period with consistent measurement of multi-generational transmission of economic characteristics, with some notable exceptions that use administrative data across several generations. Lindahl *et al.* (2015) combine a sample from the city of Malmö, Sweden in the 1930s with later administrative data and find evidence of persistence in education and income across generations. Dribe & Helgertz (2016) use data from five rural parishes in southern Sweden and observe persistence in occupational status, but not income, with no substantial changes over time. Ferrie *et al.* (2016) find some evidence of multigenerational educational persistence using U.S. census data from 1910 onward but suggest that this could be spurious due to challenges in measuring completed education precisely. Knigge (2016) uses marriage registers from five Dutch provinces and finds evidence of a moderate influence by grandfathers on occupational status which is constant over the period studied (the nineteenth and early twentieth century).[1] The present study also contributes to a broader literature on multigenerational persistence based on survey data or pseudo-linked panels.[2] The results presented here indicate that there is more intergenerational persistence in economic outcomes when grandfathers' occupations are included in addition to fathers' occupations.

This paper is structured in the following way: First, the remainder of this introduction describes a simple theoretical framework. Section 2 presents the data and discusses linkage and the economic development that took place in the period covered by this study. The main analysis using four occupational categories is conducted in Section 3. Section 4 demonstrates

[1] When data is drawn from a limited geographical region, results could be biased, as those who migrate into (in the first and third case above) and out of the region (second and third case) are not covered. Moreover, smaller regions could have particular characteristics in terms of industrial structure or demography that limits the way results can be generalized into countries as a whole. This is less of a problem in the U.S. study, which is presumably representative of the nation as a whole (but has low match rates). In any case, the present paper is the first to use data drawn from an entire country and covering three centuries (a measurement span of 100 years for each generation).

[2] A partial overview of the literature is given by Solon (2018). Examples of survey-based studies are Chan & Boliver (2013) and Hertel & Groh-Samberg (2014), who find some grandparental effects on social class; Warren & Hauser (1997), who use several composite outcomes but find no evidence; and Zeng & Xie (2014), Braun & Stuhler (2018) and Kroeger & Thompson (2015) who find some evidence of persistence in education. Only the latter two have some coverage of the time dimension in that more than one survey from the same country is utilized. Lindahl *et al.* (2015) is also partly based on survey data, augmented with administrative registers. A brief discussion on the relationship between survey-based and register-based studies of multigenerational persistence is given in the Online Appendix, Section A.1.

Studies based on pseudo-links use administrative data, but without direct linkage between individuals. Rather, information on the joint distribution of names and economic outcomes is utilized (Clark & Cummins, 2015; Clark, 2014; Olivetti *et al.*, 2018; Guell *et al.*, 2015). Interpretation of these results is sensitive to the distributions of surnames or estimation of specific parameters, making comparisons with conventional measures of intergenerational mobility/persistence challenging.

that multigenerational persistence is also found when more detailed measures of occupational status are used. Section 5 repeats the analysis on subsamples in which grandfather and grandson had less opportunity to interact directly and finds that there is high persistence also in this case. Finally, Section 6 provides a conclusion.

1.1 Theoretical framework

To fix ideas of how to interpret multigenerational persistence, a common starting point is the standard latent factor model. The latent factors are heritable across generations, and influence economic outcomes. Either relationship — the heritability across generations or the relationship between the latent factor and outcomes — could change over time. This model is operationalized by Braun & Stuhler (2018) in two equations:

$$y_{i,t} = \rho e_{i,t} + u_{i,t} \tag{1a}$$

$$e_{i,t} = \lambda e_{i,t-1} + v_{i,t} \tag{1b}$$

where y is the observed outcome for individual i in generation t, e is the latent factor and u, v are noise terms uncorrelated with each other and with observables. The "true" heritability of the unobserved (latent) trait is captured by λ, while correlations of observed traits across generations are also influenced by the relationship between the latent factor and economic outcomes, modelled by the transferability coefficient ρ.

It follows that there is a non-trivial relationship between a two-generation correlation and how differences in economic outcomes are transmitted across more than two generations. A strong two-generation correlation could reflect a strong transferability of the latent factor (high λ), a strong association between the latent ability factor and economic outcomes (high ρ) or both.

If the transmission processes did not change over time, one could in principle attempt to identify the parameters of (1) separately by comparing regression coefficients from two- and three-generation regressions (see e.g. Braun & Stuhler, 2018, p. 583-585). Given the large economic and societal changes throughout the nineteenth and twentieth century it is no surprise that previous studies have found substantial changes over time in the way transmission processes operate, making it challenging to infer parameters directly by comparing regression results.

An alternative interpretation of multigenerational persistence is that individuals are directly influenced by the presence of or social interaction with grandparents. A simple formulation of this mechanism can be given as

$$y_{i,t} = \gamma_1 y_{i,t-1} + \gamma_2 y_{i,t-2} + v_{i,t} \tag{2}$$

It is evident that a significantly nonzero grandparental coefficient — a positive association between grandfather and grandchild even when parental characteristics is controlled for — can indicate both a latent transferability of ability (Equations 1a-1b) or a direct transmission of characteristics from grandparents to grandchildren through personal interaction (Equation 2). This "duality" — noted both by Braun & Stuhler (2018) and earlier work such as Mare (2011) — shows that multigenerational persistence does not in itself point to one explanatory model being more "correct" than the other.

As has been the convention in the economics literature, the models specified by (1-2) are formulated for continuous outcomes such as income. Most existing studies on multigenerational persistence that use administrative data (e.g. Lindahl *et al.*, 2015; Ferrie *et al.*, 2016) use continuous status variables. However, in the present study, occupational data are primarily treated as discrete and unordered than continuous and/or ordered.[3] Ordered rankings of occupations become harder to interpret when comparisons are made across a long period, as it is difficult to take into account changes in relative status and/or payoff over time. For this reason, the present paper uses a measure of multigenerational occupational persistence that is not dependent on any particular ordering. While approaches to multigenerational processes in unordered outcomes have been discussed in the theoretical literature (see for example Hodge, 1966), these issues have previously not been taken into account in empirical studies of multigenerational persistence.[4]

Before presenting the analysis, we now turn to a description of the data and institutional context of the study.

2 Data and economic context

2.1 Sample construction

All the data used in this paper is obtained from official statistical sources. Full-count census data with information on all individuals in Norway, including occupation and location of residence, are available for the years 1865, 1900, 1910, 1960, 1970, 1980 and 2011. From 1960, all information on individuals can be linked using the Norwegian national ID number. Individual records from before 1960 are linked using name, birth time and birth place.[5] As linking women whose last

[3] Long & Ferrie (2018) impute incomes for five occupational groups and use OLS regressions (some categorical matrix comparison is done, but no interpretation of the magnitudes is provided), while Dribe & Helgertz (2016) use an ordered logit model based on status rankings. Knigge (2016) uses a different approach based on occupational status in multilevel models.

[4] The study of two-generation (father-son) mobility by Long & Ferrie (2013a) takes into account changes in occupational distribution, but the main results present an aggregation of odds ratios (the Altham statistic, see also the Online Appendix, Section B.1) that is not as easy to interpret. The discussion of how to take changing marginal distributions into account was pursued further in published comments to Long & Ferrie (2013a) (Xie & Killewald, 2013; Hout & Guest, 2013; Long & Ferrie, 2013b).

[5] The use of linked historical micro data in economic studies has become more widespread in the last decade; an overview is given in two recent working papers (Bailey *et al.*, 2019; Abramitzky *et al.*, 2020). The present approach has most in common with what Bailey *et al.* denote "method D", where potential multiple matches are disambiguated based on a scoring system that weighs differences in characteristics against each other. Moreover, the method used here takes into account possible conflicting matches in both data sets, in principle evaluating dissimilarity scores for all combinations of observations from period 1 and period 2 (cutoffs in the algorithm limit the number of calculations that have to be done in practice).

names change at marriage presents major challenges, and there is little occupational information on women prior to 1960, this study will focus mainly on men and paternal lineages.

We construct samples by selecting a set of birth cohorts for the "son" population to achieve links that are as comprehensive as possible. For this to be achieved, three conditions must be fulfilled. First, there must be good father-son links within each source to connect generations together. Second, for the time periods before 1960, individuals must be linked between two different sources (census records). Third, the children, fathers and grandfathers must be of working age in the years when economic characteristics can be observed.

The unit of observation in this study is a dynasty of three generations — son, father and grandfather. The construction starts with the "son" observation, observed as an adult. Each son is identified when young, using either the population registry links (1960 and onward) or name and place and date/year of birth (before 1960). Then the father of that young individual is located, and his occupation used as the "father" observation. Finally, this father is again observed as a child in a third source, and his father's occupation used as the "grandfather" observation.

The spacing of the censuses is used to group the observations into four distinct samples from four periods. Table 1 provides an overview of the size and data coverage of these samples. The earliest sample ("A") observes grandfathers in 1865, fathers in 1900 and sons in 1910, while the final sample ("D") observes grandfathers in 1960, fathers in 1980 and sons in 2011. The sizes of the samples range from 2,086 lineages in Sample A to 131,194 lineages in Sample D. For the two final samples it is also possible to add more ancestors; we return to this in Section 3.4 below.

Occupation is the only variable that is recorded throughout the period. Occupations are grouped into four major categories that are frequently used in the analysis of intergenerational mobility (Long & Ferrie, 2013a; Boberg-Fazlic & Sharp, 2018; Azam, 2015): White-collar, Farmer, Manual skilled and Manual unskilled. To avoid life-cycle bias, only occupational information on individuals between age 30 and 60 is used. Because of the long time period covered with associated changes in the relative status of occupations, the baseline specification does not use imputation of status or income level by occupation.[6] We return to imputation of economic status to occupations in Section 4.1.

For each of the four samples, we group individuals in three generations (grandfather, father, son) into four occupational categories (White collar, Farmer, Manual skilled, Manual unskilled), giving a $4 \times 4 \times 4$ matrix of occupational attainment. The number of individuals in each of the 64 cells is shown in Table A1.

A preview of the shape of persistence for white-collar occupations over time in Norway is

In their review of recent practices, Abramitzky *et al.* (2020) express a high level of confidence in the use of automatic linkage methods to compare economic mobility across different countries. In some ways, the Norwegian data differs from the U.S. data that is typically discussed in overviews of linking methods (such as the two papers referred above). Most importantly, birthplaces in Norway are reported on the municipality level, as opposed to the state level in the United States, giving a much higher resolution of an important characteristic used in matching individuals. The linkage procedure is documented in further detail in the Online Appendix, Section A.1.

[6] Censuses before 1960 do not list education, and income data is available electronically only from 1967 onward.

Table 1: Overview of data set

	(1)	(2)	(3)	(4)
Sample	A	B	C	D
	1865-1910	1865-1960	1910-1980	1960-2011
Sample size (number of lineages)	2,086	6,040	28,091	131,194
Son (index generation)				
Occupation observed in year	1910	1960	1980	2011
Birth year range	1870-1880	1900-1910	1920-1950	1960-1981
Median birth year	1878	1904	1943	1973
Number of distinct individuals	2,086	6,040	28,091	131,194
Known labor income (age 28-32)			23,775	131,094
Known labor income (age 35-39)			27,456	114,735
Known total income (age 59-63)		2,892	26,748	
Known total income (age 63-67)		5,293	25,999	
Father				
Occupation observed in year	1900	1910	1960	1980
Birth year range	1840-1865	1850-1880	1900-1910	1921-1950
Median birth year	1847	1861	1905	1945
Number of distinct individuals	1,933	4,660	21,838	95,652
Known labor income (age 28-32)				124,391
Known labor income (age 35-39)				130,644
Known total income (age 59-63)			18,177	125,567
Known total income (age 63-67)			25,587	122,639
Grandfather				
Occupation observed in year	1865	1865	1910	1960
Birth year range	1805-1835	1805-1835	1850-1880	1900-1930
Median birth year	1815	1824	1870	1912
Number of distinct individuals	1,893	4,529	19,702	84,292
Known total income (age 59-63)				107,764
Known total income (age 63-67)				115,628
Great-grandfather				
Occupation observed in year			1865	1910
Birth year range			1805-1835	1850-1880
Median birth year			1825	1870
Sample size (number of lineages)			2,422	19,700
Number of distinct individuals			1,668	11,468
Great-great-grandfather				
Occupation observed in year				1865
Birth year range				1805-1835
Median birth year				1824
Sample size (number of lineages)				1,676
Number of distinct individuals				967

Note: Years of observation, birth years and number of observations in each of the four samples. Men in Norway, occupations observed in national census

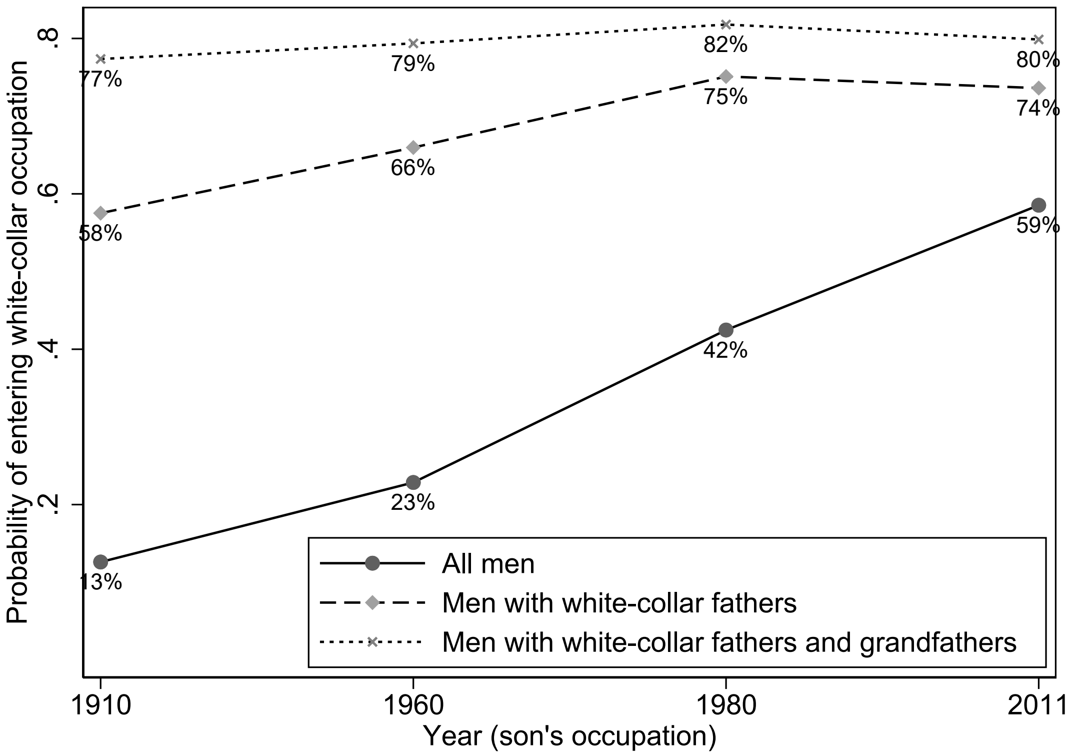

Figure 1: Probability of entering a white-collar occupation, by family background (father's occupation). Men in Norway, selected cohorts.

given in Figure 1. The lowermost line shows the overall probability of entering a white-collar occupation for men, measured as the share of individuals (subject to some sample restrictions described below) with such occupations at each of four censuses. In 1910, 13% of the men in the sample held a white-collar occupation while the share in 2011 is 59%. The middle line shows similar probabilities for men whose fathers also held white-collar occupations. Despite the low rates in the overall population in 1910, a majority of sons of white-collar men are able to enter white-collar occupations, reflecting limited intergenerational mobility. The uppermost line, however, shows that the importance of family background extends beyond father-son associations. If the grandfather also held a white-collar occupation, the probability in 1910 jumps from 58% to 77% and in 2011 from 74% to 80%.

2.2 Representativity of the sample

The linkage of individuals on name, time of birth and place of birth means that the selection into the sample used in this paper is not completely random. While great care has been taken to link individuals by means of time-invariant characteristics only — name, birth time and birth place — the structure of the censuses used as base data means that only observations on families with parent-child age differences that match the observation periods can be used for linkage. Moreover, individuals whose characteristics are less unique (common names, born in large municipalities) are harder to link in a reliable manner than those with less common names from small places. This section gives an overview of how some of these challenges are handled.[7]

When links are made across three periods (grandfather-father and father-son) the match rates compound. In sample B, for example, in the population for which we know the son's occupation in 1960, in 36.1% of cases we have a 1910 father's observation that satisfies all the criteria: that the son is identified in 1910; that there is a link between son and father in 1910; that the father is in the correct age interval (30 to 60) and that the father has a recorded occupation. Among these father-son pairs we can identify 8.5% of the grandfathers in 1865 according to the same criteria, giving a final sample size of 3.5% of men born between 1900 and 1910. The corresponding gross match rate — the share of individuals whose occupations are observed at t for whom both fathers and grandfathers are known, with an observed occupation, and between 30 and 60 years old at $t-1$ and $t-2$ respectively — for sample A is 1.4%, for sample C 4.4% and for sample D 24.9%.

Because the selection process may change over time, we should as a rule be careful in interpreting small differences between periods as time trends. However, as the main purpose of this paper is to document the robustness of the economic impact of grandfathers over and above that of fathers in different periods and for different economic variables, biases that vary in magnitude over time do not necessarily weaken the analysis if the effects of these biases occur in all of the samples.

Any biases in the samples are handled in the analysis in three ways. First, the estimates of

[7] The relationship between the way records are matched and the results on multigenerational persistence will be further discussed below in Section 3.3.

persistence are based on odds ratios, which are invariant to changes in the marginal distribution. This means that over-representation of a given occupational group does not directly drive the estimation results. Second, age controls for all generations are added to all occupational regressions. Third, individuals (in the final generation) born in 1881-1899, 1911-1919 and 1951-1959 are excluded from the analysis (as shown in Table 1) because their ancestors' year of birth fits poorly with the years in which occupations and family links can be observed.

2.3 Economic development and intergenerational mobility in Norway

In 1865, Norway was a predominantly rural society; 40 % of the adult male population were farmers (owners, tenants or managers), while an additional 20 % were cottagers with limited property rights. The oldest grandfathers in this study were born in 1815, immediately after the end of the Napoleonic wars and contemporaneous hunger and economic crisis in Norway. From 1860 to 1913 there was substantial emigration to the United States, with more than 800,000 individuals emigrating (Norway's total population in 1865 was 1.7 million).

Norway industrialized relatively late compared with core European countries, but around the turn of the century many industrial ventures were started, often in locations dictated by the availability of hydroelectric power. In 1910, 32 % of the working-age male population were farmers and 31 % list a manual skilled occupation. During this period, Norway was heavily dependent in terms of large-scale emigration, food imports and raw material exports on the world economy, even though many people still lived on small farms in remote areas, and had to travel substantial distances to even the closest urban center.

After 1910, in which the final generation in sample A is observed, the economic development of Norway shared several characteristics with the rest of western Europe. While Norway was neutral during World War I, the economy was still affected, with increasing prices causing hardship for the poor and high shipping rates profiting a small group of shipowners. The first decades of the twentieth century represented a period of increasing power for the labor unions, with the first stable Labor Party government being formed in 1935. The country was under German occupation from 1940 to 1945, though material destruction was limited except in the far North. The 1950s and 1960s saw rapid economic growth, and the number of workers in manufacturing peaked in this period. This is widely regarded as an era of equalization of opportunities, with the quality of elementary education improving. Aaberge *et al.* (2020) find that income inequality in Norway was relatively high until the late 1930s but fell to lower levels by the early 1950s.

Semmingsen (1954) ties the emergence of the Norwegian industrial and middle classes from the 1860s onwards to the large population movements in the second half of the eighteenth century. The grandfathers in samples A and B are hence observed as adults in 1865 in a predominantly agricultural society with relatively low social fluidity. While the father generation had more economic opportunities in term of industrial employment, Modalsli (2017) documents that father-son occupational mobility in Norway in 1865-1900 period was low compared to contemporary United States and twentieth century Norway.

Pekkarinen *et al.* (2017) find that intergenerational mobility (measured by brother as well as father-son income rank correlations) increased from the 1950s onwards, with lower correlations for children born after the Second World War. This is a period of increased spending on primary education, as well as several expansions of social insurance and other social programs.

Since the start of North Sea oil production in the 1970s, economic growth in Norway has continued at a fast pace, with Norwegian GDP per capita ranked as one of the highest in the world. The labor force is increasingly concentrated in white-collar occupations. While Modalsli (2017) finds an increase in father-son occupational mobility until the 1980-2011 period, estimates based on intergenerational income elasticities (for example Bratberg *et al.*, 2005; Nilsen *et al.*, 2012) find some evidence of decreasing parent-child income mobility around the turn of the twenty-first century.

3 Multigenerational occupational persistence

With the data structured as described in the previous section, where each observation consists of information on economic status from three generations, we now turn to the measurement of persistence using occupational categories.

3.1 Odds ratios across occupations

Transmission of discrete characteristics such as occupation groups across generations are best understood using transition matrices. Consider the parent-child transition probability matrix

$$P = \begin{pmatrix} p_{ii} & p_{ij} \\ p_{ji} & p_{jj} \end{pmatrix} \tag{3}$$

where the parent's occupation is held constant across rows and the child's occupation across columns. For example, if we denote white-collar occupations by i and non-white-collar occupations by j, the variable p_{ii} states the probability that the child of a white-collar parent obtains a white-collar occupation, while p_{jj} is the probability that the child of a non-white-collar parent obtains a non-white-collar occupation.

In this simple case of two generations with two occupational groups for each generation, we can measure intergenerational mobility using a two-way odds ratio (Agresti, 2002, p. 44). Denoting the probability of the son of a father with occupation i entering an occupation j as p_{ij}, the odds ratio for father-son mobility is

$$\Theta = \frac{p_{ii}/p_{ij}}{p_{ji}/p_{jj}} \tag{4}$$

A high value of Θ corresponds to high intergenerational persistence (that is, low intergenerational mobility), while a value of 1 can be interpreted as no association between father's and son's occupations (no intergenerational persistence and very high intergenerational mobility).[8] One advantage of using odds ratios rather than simple transition probabilities is that we abstract from changes in the marginal distribution of occupations. For example, consider i as a white-collar occupation category with only 10 per cent of the population. In this case, the probability that the son of a white-collar father obtains a white-collar job is is 17 per cent (and 83 per cent that he does not, giving an odds of $0.17/0.83 = 0.20$). The son of a non-white-collar father has a 9 per cent chance of getting a white collar job (and 91 per cent that he does not, odds of $0.09/0.91 = 0.10$). Θ is the ratio of these odds. In this case, $\Theta = 2(0.20/0.10)$, a relatively high persistence rate.

For each occupational category and time period, we can create 2×2 tables indicating whether fathers and sons hold the relevant occupation or not and calculate odds (probability ratios) and odds ratios (Θ). As shown in Table 2, for white-collar occupations, we obtain very high odds ratios Θ in the early period, of 15.5 for sample A and 9.7 for sample B. An odds ratio of 9.7 means that the probability ratio (odds) of the son of a white-collar father entering a white-collar occupation is 9.7 times higher than the corresponding ratio for a son of a non-white-collar father. For sample C (1910-1980) the odds ratio is 6.3 and for sample D (1960-2011) it is 3.1, reflecting increased mobility. For farmers there is no such trend towards mobility. Skilled workers have initially higher persistence but follow a similar trend to that of white-collar workers, while the trend for unskilled workers is less clear.

One can also calculate odds ratios for grandfather-grandson tables in a similar manner, as shown in Table 2. These persistence measures are slightly lower than the father-son ratios.[9] However, such associations do not incorporate the information from the father generation. For this reason, we now move to a framework where we can utilize information from all three generations.

3.2 Binary outcomes across three generations

The simplest way to construct a three-generation analogy of two-generation odds ratios is to use a logit model. We choose an occupational category and set the outcome variable to 1 if this occupation is entered by the final generation and 0 if it is not entered. The baseline approach in the present paper is then to regress this outcome against father's characteristics X^f and grandfathers' characteristics X^g as background variables (ι indexes the dynasty):

[8] Values of less than 1 indicate that families within the occupation group are *less* likely to transition into the group than individuals outside. Values substantially below 1 are not often seen empirically, and such outcomes (which are analogous to negative intergenerational income elasticities) are rarely discussed in the intergenerational mobility literature.

[9] Associations with standard errors and controls for age are given in Columns (1) (grandfathers) and (7) (fathers) in Tables 6 and A3-A5.

Table 2: Odds ratios calculated from 2×2 tables, for four occupational classifications.

	(1)	(2)	(3)	(4)
Sample	A 1865-1910	B 1865-1960	C 1910-1980	D 1960-2011
Occupation: White collar				
Odds ratio Θ (Father-son)	15.5	9.7	6.3	3.1
Odds ratio Θ (Grandfather-grandson)	11.2	6.6	3.8	2.4
Share of sons in occupation category	13 %	23 %	42 %	59 %
Occupation: Farmer				
Odds ratio Θ (Father-son)	4.4	9.3	24.4	20.3
Odds ratio Θ (Grandfather-grandson)	2.8	2.9	6.4	10.2
Share of sons in occupation category	22 %	28 %	10 %	2 %
Occupation: Manual, skilled				
Odds ratio Θ (Father-son)	6.4	3.4	2.3	2.2
Odds ratio Θ (Grandfather-grandson)	3.6	1.9	1.2	1.2
Share of sons in occupation category	24 %	37 %	41 %	31 %
Occupation: Manual, unskilled				
Odds ratio Θ (Father-son)	1.6	2.5	6.3	2.6
Odds ratio Θ (Grandfather-grandson)	1.1	1.5	2.4	1.9
Share of sons in occupation category	42 %	13 %	6 %	9 %

Note: Higher odds ratios indicate higher persistence.

$$\log\left(\frac{\Pr(\text{Son's occ} = Z)_\iota}{\Pr(\text{Son's occ} \neq Z)_\iota}\right) = \alpha + \beta \boldsymbol{X}^f_\iota + \gamma \boldsymbol{X}^g_\iota + \sum_{q \in \{s,f,g\}} \left(\delta \cdot age^q_\iota + \zeta \cdot (age^q_\iota)^2\right) + \epsilon_\iota \tag{5}$$

In the case where there is no grandparental information ($\boldsymbol{X}^g$ is empty), fathers' characteristics are represented by a simple 0-1 dummy for occupational category and there are no age controls for the son generation, the estimates of β from Equation (5) are equivalent to the log of the odds ratios from the 2×2 table.

The setup in Equation (5) will form the basis of the analysis of occupational persistence in this paper. Isolating persistence in this way is important, as most countries have experienced major changes in both the distribution and the income rank of occupations over time. In particular, the number of farmers, a heterogeneous group that cannot always be ranked reliably relative to non-farming occupations, was very high in many countries in the late nineteenth and early twentieth century, and measures based exclusively on incomes or ranked occupations could hence give misleading results when applied over such a long time range.

The simplest joint model of fathers and grandfathers uses a dummy variable D for each generation that is equal to 1 if that generation holds the occupation in question. This will be the baseline specification. Age controls will also be used throughout. Hence, we have the following expression for the covariate vector for generation $q \in (f, g)$ in Equation (5):

$$\boldsymbol{X}_\iota^q = D_\iota^q \tag{6}$$

i.e. the model

$$\log\left(\frac{\Pr(\text{Son's occ} = Z)_\iota}{\Pr(\text{Son's occ} \neq Z)_\iota}\right) = \alpha + \beta D_\iota^f + \gamma D_\iota^g + \sum_{q\in\{s,f,g\}} \left(\delta \cdot age_\iota^q + \zeta \cdot (age_\iota^q)^2\right) + \epsilon_\iota \tag{7}$$

is estimated four times, with the indicator variable D as White collar, Farmer, Manual skilled and Manual unskilled, respectively. The resulting parameter estimates are shown in Table 3. In line with the common terminology of changes in economic characteristics across generations, persistence and mobility will be taken as opposites: high mobility equals low persistence and vice versa.[10]

We start with the outcome of the son entering a white-collar occupation, as opposed to entering an occupation in one of the other three categories. The corresponding right-hand-side variables are a dummy variable for whether the father had a white-collar occupation, a dummy for whether the grandfather had a white-collar occupation, and second-degree polynomials controlling for the age (at the time of observation) of each of the three generations.[11] The top panel of Table 3 shows the exponentiated coefficients for father's and grandfather's occupations and can be interpreted as odds ratios.

The coefficient on father's occupation — 11.8 — is similar in magnitude to that in the two-generation case reported in the previous section. We now focus on the coefficient on grandfather's occupation. For an individual observed in 1910 (sample A) with a given father's occupation, having a grandfather with a white-collar occupation increases the odds of entering a white-collar occupation by 2.8. In other words, the grandfather effect in the 1865-1910 sample is comparable in size to the father effect in the 1960-2011 sample. The grandparental effect in sample B (sons observed in 1960) is also large, while the grandparental coefficients in samples C and D (sons observed 1980 and 2011) are lower, in accordance with the generally higher mobility into and out of white-collar occupations. However, all coefficients are significant and substantial.

The second panel of Table 3 reports father and grandfather coefficients for farmers, where again the variables of interest are dummies for whether the father and grandfather belonged to the same occupational category. Once again, there are large and highly significant coefficients. For both white-collar occupations and farmers, the grandparental coefficient in all four samples is significant at the 1% level. The grandfather odds ratio for farmers is moderate in the two

[10] In some cases, interpreting logit coefficients across samples can be misleading (Mood, 2010). However, because the marginal distributions of occupations change over time, some normalization of marginal distributions is necessary (see the works cited in Footnote 4). Moreover, as argued by Buis (2016), there are some applications (such as the importance of social backgrounds) where direct interpretation is appropriate. For this reason, the logit model will be used as a baseline specification. Results using linear probability models are available on request.

[11] Age controls do not matter greatly for the results.

Table 3: Odds ratio coefficients for binary occupational regressions on father's and grandfather's occupations

	(1)	(2)	(3)	(4)
Sample	A	B	C	D
	1865-1910	1865-1960	1910-1980	1960-2011
Occupation: White collar				
Father	11.79***	8.071***	5.151***	2.730***
	(13.58)	(23.28)	(48.16)	(79.46)
Grandfather	2.838***	2.504***	1.802***	1.631***
	(3.51)	(6.19)	(13.93)	(30.26)
Occupation: Farmer				
Father	3.686***	8.179***	18.71***	8.636***
	(8.31)	(24.49)	(44.78)	(43.94)
Grandfather	1.595***	1.471***	1.929***	3.916***
	(2.96)	(5.10)	(10.84)	(25.13)
Occupation: Manual, skilled				
Father	5.458***	3.312***	2.351***	2.171***
	(12.26)	(17.07)	(31.97)	(62.26)
Grandfather	2.039***	1.316***	0.959	1.021*
	(3.55)	(2.84)	(-1.38)	(1.68)
Occupation: Manual, unskilled				
Father	1.688***	2.327***	5.581***	2.210***
	(3.80)	(8.32)	(30.68)	(23.37)
Grandfather	0.935	1.223**	1.554***	1.652***
	(-0.57)	(2.25)	(7.06)	(18.73)
N	2086	6040	28091	131194
Son observed:	1910	1960	1980	2011
Father observed:	1900	1910	1960	1980
Grandfather observed:	1865	1865	1910	1960

Exponentiated coefficients; t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Higher coefficients indicate higher multigenerational persistence. Separate logit regressions for each sample and occupational category. Dependent variable: dummy for son's occupation. Constant terms and coefficients on quadratic controls for age for all three generations were also included in the regressions.

earliest periods (samples A and B) but substantially higher for the later samples (C and D), with a high of 3.9 for the final period.

While white-collar persistence is higher in early periods and farmer persistence is higher in later periods, we have a qualitatively similar finding: that grandfathers matter. Their influence declines over time, but is still apparent even in the most recent generation. Knowing that an individual had a father in the given occupation group substantially increases the probability that the individual himself holds the occupation. Controlling for father's occupation, knowing that the individual had a grandfather in this category also increases the probability substantially, but less so than for the father. This is strong evidence that the multigenerational process in these occupations is more persistent than one would expect from a simple father-son regression.[12]

The two remaining occupation groups, manual skilled and manual unskilled occupations, exhibit larger differences in regression results across samples. For manual skilled occupations, the picture is similar to that of white-collar occupations in samples A and B, with statistically significant coefficients of substantial magnitude. In sample C, the grandparental coefficient is negative; this means that when we compare two individuals with the same father's occupation, the one whose grandfather did not have a manual, skilled occupation would have a higher probability of entering such an occupation (though this difference is not statistically significant). Examining the full matrix of occupations shows that this negative coefficient is a result of the relative probabilities for grandsons of manual unskilled individuals, who are more likely to enter manual skilled occupations than any other group. For manual unskilled occupations, the coefficient on grandfathers is significant in samples B-D, while it is close to zero and insignificant in the initial period.[13]

Just as the magnitude of the observed associations differs depending on which occupational category one uses to split the population, the evolution of multigenerational persistence over time also differs across occupation categories. While there is a clear evolution towards lower persistence for white-collar and manual skilled occupations, persistence has increased for farmers and manual unskilled occupations. One way of interpreting these differences is that decreasing father-son persistence (increasing mobility) occurs together with lower persistence also for earlier ancestor generations.

The significance of grandfather's occupation does not depend on the specific grouping of occupations used here. Table A2 in the Online Appendix shows coefficients from estimations on smaller occupational groups. In the case of specific occupations with a limited number of individuals, some of the cells in a $2 \times 2 \times 2$ transition matrix will frequently not be fully occupied — in these cases, coefficients cannot be estimated. However, where there are a sufficient number of observations, the pattern for the detailed sample occupations is similar to that in Table 3, though slightly higher on average (as would be expected from more precise categories).

[12] An explicit comparison of the models, including measures of model fit, is presented in Section 4.2 below.

[13] One could also add an interaction term between grandfather's and father's occupation. The coefficient on such an interaction term is in most cases close to 1 (no effect) and insignificant, and does not substantially alter the coefficent reported for grandfather's occupation here.

The way persistence is modelled here — considering one occupational category at a time rather than all four categories jointly — could potentially hide some characteristics of multigenerational persistence that only emerge when the full $4 \times 4 \times 4$ matrix is considered. For example, the probability of a son entering a white-collar occupation could differ depending on whether the father had a manual skilled or manual unskilled occupation. However, careful consideration of the odds ratios from such interactions reveals no substantial interactions that change the interpretations above. Including them in a measure of average odds ratios (the Altham statistic) confirms a picture of reduced grandparental influence over time. This is laid out in detail in the Online Appendix, Section B.1.

Finally, the categories used here are based on the characteristics of occupations and may therefore change in social status over time. This could be a concern for white-collar occupations, which today encompass a broader segment of society than 150 years ago, as the size of the occupation group has increased. For this reason, in Section 4 we examine status-based measures of multigenerational persistence and show that splitting the sample by high- and low-status occupations yields a similar result to that found for white-collar occupations. On the other hand, the changing role of persistence in farming over time is possibly better understood with respect to structural change and the decreasing number of farms. Though the first-order decrease in occupation size is implicitly controlled for in the construction of the odds ratio, it could still be the case that the remaining population of farmers is in some way "selected", so that persistence is higher among the families that have remained in this occupational category.

3.3 Record matching and the measurement of persistence

To assess to what extent the results obtained in this paper can be regarded as valid for the entire population, we can compare father-son intergenerational mobility for the subsample whose grandfather's identity is unknown with those for whom a father-grandfather link is successfully obtained. The difference between odds ratios when the same controls for age are imposed on both samples is given in Table 4. The table reports the regression results for a dummy variable for whether grandparent's occupation is available, interacted with father's occupation in a regression where the outcome variable is son's occupation. This could be interpreted as the ratio of odds ratios calculated on the matched versus the unmatched sample.

The coefficients are exponentiated; a number larger than 1 means that father-son pairs for which the grandfather is known ("matched dynasties") have higher measured persistence than father-son pairs with an unknown grandfather ("unmatched dynasties"). For example, in sample A, the matched dynasties have 0.2 per cent higher persistence than unmatched dynasties, while in sample B, the matched dynasties have 22.9 per cent higher persistence.

In general, these differences are small, indicating that the findings are not driven by artifacts of the matching process. However, in particular for farmers and manual unskilled workers in the first sample there are larger differences, with matched dynasties experiencing lower persistence than unmatched dynasties. The results for these groups in these time periods should therefore be interpreted with some caution.

Table 4: Interaction effects between father's occupation and whether father-son pair can be linked to a grandfather.

	(1)	(2)	(3)	(4)
Sample	A	B	C	D
	1865-1910	1865-1960	1910-1980	1960-2011
Occupation: White collar				
Interaction term	1.002	1.229**	1.043	1.018
	(0.01)	(2.33)	(1.03)	(1.20)
Occupation: Farmer				
Interaction term	0.693**	0.958	1.079	0.830***
	(-2.34)	(-0.50)	(0.98)	(-3.58)
Occupation: Manual, skilled				
Interaction term	0.963	1.031	1.080**	1.005
	(-0.26)	(0.44)	(2.41)	(0.34)
Occupation: Manual, unskilled				
Interaction term	0.634***	0.809**	1.128*	1.178***
	(-3.20)	(-2.09)	(1.81)	(4.04)
N	10238	71156	76825	388502
Son observed:	1910	1960	1980	2011
Father observed:	1900	1910	1960	1980
Grandfather observed:	1865	1865	1910	1960

Exponentiated coefficients; t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Outcome variable is dummy for son's occupation; other controls are age, father's occupation and linkage status separately.

An alternative approach to this direct comparison of samples is to weight observations by the match rate of subgroups, as suggested by Bailey *et al.* (2020). Here, we follow this approach as implemented in Ferrie *et al.* (2020) and regress the probability of being in the sample on a set of covariates, using this to predict a matching probability at the individual level. This information is then used in an inverse probability weighting framework to re-run the persistence regressions on the weighted sample. The full table of weighted and unweighted regression results are given in the Online Appendix, Table A7. For three of the samples (A, C and D) there are only minor differences between the weighted and unweighted analysis. For sample B the differences are larger, for both father and grandfather coefficients. However, there does not appear to be a systematic bias, as coefficients are in some case higher (for example for father and grandfather farmers) and in other cases lower (for example for manual unskilled grandfathers). The large deviations for sample B occur because a few cells (combinations of occupations across generations) have very few observations, and these cells are given very high weight and undue influence in the weighted regressions.[14]

[14]For the final sample, results have also been replicated on subgroups differentiated by age. This does not substantially alter the results; see the Online Appendix, Section A.5.

3.4 What about the great-grandfathers?

So far, we have examined the influence of two generations of ancestors on the outcome of the final generation. The data allow examination of the influence of an even larger set of generations. Two caveats must be kept in mind when conducting such an analysis. First, as match rates are imperfect, the reduction of the sample size is compounded when more generations are considered and the observation years are irregularly spaced. Second, the number of ancestors increases geometrically with the number of generations, and we only consider paternal ancestors here.

Table 5 shows the results of logit regression with son's occupation as the outcome for models with further ancestor generations for each of the four occupational groups. In sample C, information on the great-grandfather is available; in sample D, we have information on both great-grandfathers and great-great-grandfathers.

The first column of Table 5 shows the results of the four-generation models, with the paternal lineage observed in 1865, 1910, 1960 and 1980, respectively. For white-collar occupations, great-grandfathers have substantial predictive power, with estimated 53% higher odds of entering a white-collar occupation for the great-grandson of a white-collar worker conditional on father's and grandfather's occupations. There are substantial coefficient values also for farmers and unskilled manual workers, though these are not statistically significant. Table A6 compares three-generation regressions for the baseline sample and for the subsample where the fourth generation is available; when the number of observations is reduced from 28,091 (with three generations) to 2,422 (with four generations) several of the grandparental coefficients lose significance.

The second column of Table 5 shows the same model as the first, with estimated coefficients for four generations, but in a later time period (measured in 1910, 1960, 1980 and 2011). We again observe a statistically significant positive coefficient for the great-grandfather for white-collar occupations, and now also for farmers. In the third column, a fifth generation (great-great-grandfathers) is also introduced. This reduces the sample size substantially, and no great-great-grandfather coefficients are significantly different from zero. For manual skilled occupations, we observe negative coefficients for great-grandparents in all three models (statistically significant in the second model). For given father's and grandfather's occupations, an individual would have a *lower* probability of entering a manual skilled occupation if his grandfather was a manual skilled worker. This is again driven by the higher probability of descendants of manual unskilled workers of entering manual skilled occupations.

4 Persistence and the measurement of economic characteristics

While the analysis above shows positive and significant coefficients for grandfathers in all time periods and for most occupational categories, one might fear that this effect was driven by changes in the status of occupations, or by otherwise coarse measurement of the father characteristics.

Table 5: Odds ratio coefficients for binary occupational regressions on four and five generations

	(1)	(2)	(3)
Sample	C	D	D
	1865-1980	1910-2011	1865-2011
Occupation: White collar			
Father	5.566***	2.894***	3.087***
	(14.04)	(31.51)	(9.64)
Grandfather	1.622***	1.559***	1.751***
	(2.83)	(10.05)	(3.40)
Great-grandfather	1.532*	1.185***	0.916
	(1.79)	(3.23)	(-0.40)
Great-great-grandfather			1.414
			(1.03)
Occupation: Farmer			
Father	23.15***	8.049***	6.677***
	(13.05)	(20.07)	(6.33)
Grandfather	1.426	3.873***	2.655**
	(1.62)	(9.69)	(2.54)
Great-grandfather	1.337	1.525***	1.213
	(1.62)	(3.36)	(0.47)
Great-great-grandfather			1.812
			(1.64)
Occupation: Manual, skilled			
Father	2.756***	2.222***	2.159***
	(10.71)	(24.60)	(6.97)
Grandfather	0.813	1.078**	1.198
	(-1.59)	(2.13)	(1.47)
Great-grandfather	0.897	0.810***	0.898
	(-0.61)	(-5.13)	(-0.58)
Great-great-grandfather			0.810
			(-0.86)
Occupation: Manual, unskilled			
Father	4.776***	2.138***	0.944
	(7.80)	(8.53)	(-0.14)
Grandfather	1.291	1.487***	1.151
	(1.06)	(5.21)	(0.47)
Great-grandfather	1.371	1.113	1.584*
	(1.63)	(1.45)	(1.72)
Great-great-grandfather			1.320
			(1.32)
Age controls	Yes	Yes	Yes
N	2422	19700	1676
Son observed	1980	2011	2011
Father observed:	1960	1980	1980
Grandfather observed:	1910	1960	1960
Great-grandfather observed:	1865	1910	1910
Great-great-grandfather observed:			1865

Exponentiated coefficients; t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Higher coefficients indicate higher multigenerational persistence. Separate logit regressions for each sample and occupational category. Dependent variable: dummy for son's occupation. Constant terms and coefficients on quadratic controls for age for all generations were also included in the regressions.

This section shows that the results are robust to alternative specifications. Data on incomes are used to impute status to occupations and create categories that hold status constant over time. We also establish that the grandparent coefficient remains siginficant and of a similar magnitude if more detailed occupational information on the parent generation is added.

4.1 Persistence in high- and low-ranked occupations

The analysis presented in the previous section maintained a four-way separation of occupation groups (white collar, farmer, manual skilled and unskilled). While the use of odds ratios allows for a study of changing persistence that is not mechanically driven by changes in the share of the population in each occupation category, a comparison between time periods is still complicated by these changes. Entry into white-collar occupations is much more open in the twenty-first than in the nineteenth century, and twenty-first century white-collar occupations also employ a much larger share of the population.

By shifting the focus from occupation to income, one can get more information about whether to think of the grandparental influence as horizontal movements across fields, or as vertical movements between different levels of economic status. While there is no individual-level income information available prior to the 1960s, occupational income averages can be used to impute the status of each occupation.

Average occupational incomes are obtained from official statistics. For 1980 and 2011, direct individual-level links between occupations and tax records are used to calculate occupation mean incomes. For 1960, a similar method is used, but using incomes from 1967 (the earliest available year). There are 255 distinct occupational categories in 1960, 379 in 1980 and 202 in 2011. For 1910 and 1900, data from Statistics Norway (1915) was used (based on contemporary data links between the 1910 Census and tax records from the same year), giving 107 and 103 occupational categories with distinct mean incomes. For 1865, income tabulations from Norwegian Department of Justice (1871) provide the basis for a comparable estimate with 73 categories, though with a slightly different population definition (incomes calculated on the basis of men aged 25 and above, compared to men aged 30-60 for other years).

In each year, occupations are ranked by mean income. Some occupation categories (in particular, farmers) are very large in some time periods. For this reason, it is not feasible to study arbitrary percentile cutoffs. However, for three levels there is sufficient granularity to construct reliable indicator variables for social status. At the top of the occupational status distribution, an indicator variable for being above the 11 % separates the highest-status occupations from the rest. Adding some other high-status occupation groups allows for an indicator variable for the top 22 %. At the lower end of the distribution, an indicator variable is constructed for the 22 % of the population in the lowest-paid occupations.[15]

[15] Because of the limited number of occupation categories, the size of the highest-ranked category ranges from 10.1 % (in 1960) to 12.3 % (in 1865); the wider highest-ranked category from 20.1 % (in 2011) to 23.9 % (in 1910); the lowest-ranked category from 21.9 % (in 1900) to 23.0 % (in 1980). The population shares are calculated on the full sample of men 30-60 with observed occupations.

With these constant status groupings, we maintain the setup of Equation (7) and run logit regressions for each time period. The outcome variable is an indicator variable for the son's occupational status (top 11, top 22 or bottom 22 %), while the right-hand-side variables are similar indicator variables for fathers and grandfathers, plus age controls. Figure 2 plots the coefficient on grandfather's occupational status for the four time periods. The full results are reported in the Online Appendix, Table A9.

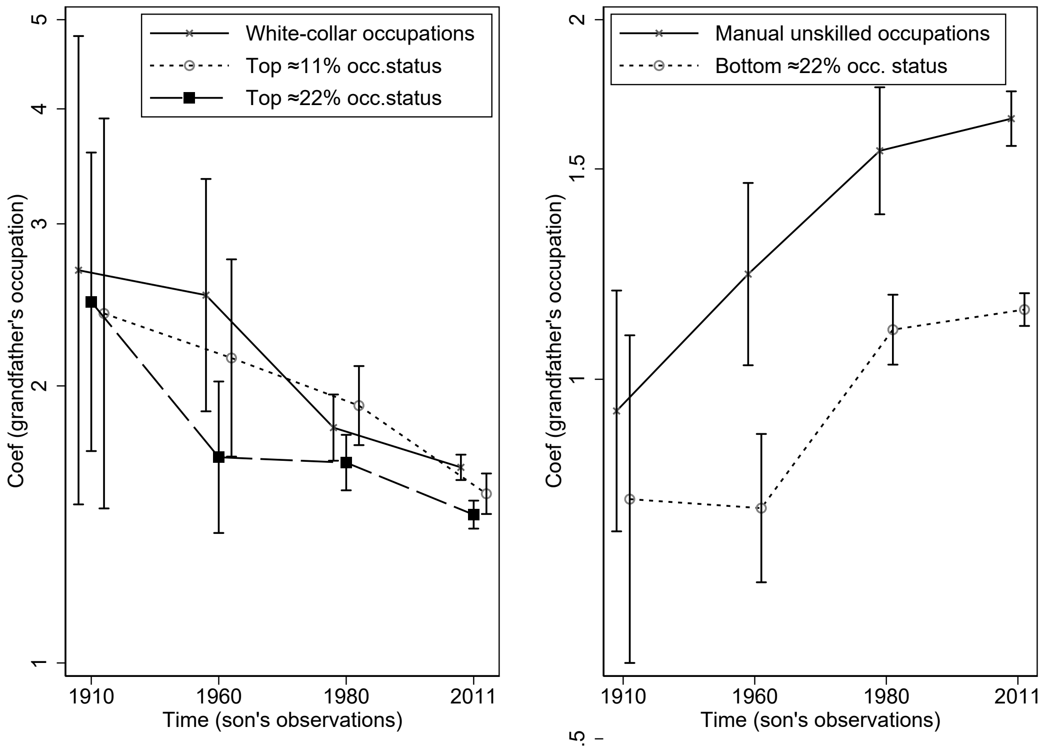

Figure 2: Changes in persistence over time for high- and low-status occupations. Figure shows exponentiated coefficient on grandfathers; separate regressions for each time period and occupation classification. 95% confidence intervals.

The left panel compares the grandfather coefficients on the high-ranked occupations to the previously reported coefficient on white-collar occupations. There is a clear downward trend (towards lower persistence) over time. Having a grandparent in an occupation category that pays at the 11th percentile or higher increases the odds of ending up in a similar category by 2.4 in 1910, 2.1 in 1960, 1.9 in 1980 and 1.5 in 2011. If we instead consider the top 22 %, the decrease is larger in the first period, but the developments in the three different ways of partitioning the population are otherwise remarkably similar.

The right panel of Figure 2 shows results for lower-status occupations. There is an increase in the grandparental coefficient for the lowest-paid occupations (the bottom 22%) over time, consistent with increased persistence for movement into and out of the category of manual unskilled occupations in general. These results show that the estimated development in multigenerational persistence over time described in Section 3 are not artifacts of changing sizes of occupational categories.

For the two final time periods the analysis of occupational status can be supplemented with results for individual incomes obtained from tax data. This analysis is presented in the Online Appendix (Section B.2). Linear regressions of income ranks of three generations provide comparable results to what we learn from occupational status.

4.2 More detailed observation of the parent generation

A relatively coarse occupation specification, with few categories, leaves open the possibility that including grandparent information simply gives us more precise information on the father's background. This could be the case for example because a high-status grandfather makes it more likely for the father to hold a high-prestige occupation within his occupational category. The reliance on paternal lines (because of limitations in the historical data) also means that the grandfather coefficient could capture the results of nonrandom marital matching, where the grandfather association is confounded with human capital transfers from the mother. This subsection shows that this is not likely to be the case.

If the results seen so far are mainly driven by imperfect measurement of the father's generation, we would expect the grandfather coefficient to be substantially reduced when more information on the middle generation is added. To investigate whether this is the case, we replace the single dummy in (5) with a more detailed specification of father's occupation (2-digit HISCO for 1900-1910; 2-digit NYK for 1960-1980). That is, while we maintain the grandfather specification $\boldsymbol{X}_\iota^g = D_\iota^g$ from Equation (6), we replace that of the father with

$$\boldsymbol{X}_\iota^f = \underbrace{\{X_\iota^{f(1)}, X_\iota^{f(2)}, \ldots, X_\iota^{f(N-1)}\}}_{\text{Dummy variables for } N \text{ occupational categories}} \tag{8}$$

Table 6 reports coefficient estimates using a range of models that incorporate different amounts of information on the parent generation. All columns report exponentiated coefficients from a logit regression where the outcome is whether the son has a white-collar occupation, similar to the first panel of Table 3. Model 1 includes only a dummy for whether the grandfather has a white-collar occupation, while Model 2 is simply a repetition of the baseline results from Section 3, where there is also a dummy for whether the father has a white-collar occupation. Comparing the two first columns, it is clear that not controlling for father's characteristics at all (Model 1) clearly ascribes too much of the family background influence to the grandfather.

In the third column, the characterization of father's occupation is extended to the full set of occupational categories used in the census data (covariates from Equation (8)). This additional

Table 6: Coefficients from regressions with more detailed information on the parent generation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sample A (1865 - 1900 - 1910)							
Father's occ		11.79***	*41 cat.*	11.55***	*41 cat.*		14.96***
		(13.58)		(10.75)			(15.96)
Mother's occ				0.44	*11 cat.*		
				(-0.84)			
Grandfather's occ	10.57***	2.84***	2.33***	5.46***	2.81**		
	(9.70)	(3.51)	(2.65)	(4.61)	(2.42)		
N	2086	2086	2041	1271	1181	–	2086
χ^2 LR	107.4	280.7	356.8	204.8	229.9		268.4
Pseudo-R^2	0.068	0.178	0.232	0.212	0.261		0.170
Sample B (1865 - 1910 - 1960)							
Father's occ		8.07***	*63 cat.*	7.60***	*63 cat.*		9.65***
		(23.28)		(21.45)			(26.46)
Mother's occ				1.67	*22 cat.*		
				(1.46)			
Grandfather's occ	6.67***	2.50***	2.32***	2.58***	2.36***		
	(14.91)	(6.19)	(5.53)	(6.08)	(5.32)		
N	6040	6040	6008	5574	5515	–	6040
χ^2 LR	266.7	824.3	933.3	719.7	843.4		785.8
Pseudo-R^2	0.041	0.127	0.146	0.120	0.143		0.121
Sample C (1910 - 1960 - 1980)							
Father's occ		5.15***	*67 cat.*	4.93***	*67 cat.*	*67 cat. + income*	6.10***
		(48.16)		(43.48)			(56.22)
Mother's occ				2.46***	*47 cat.*	*47 cat. + income*	
				(8.79)			
Grandfather's occ	3.61***	1.80***	1.53***	1.73***	1.47***	1.35***	
	(34.20)	(13.93)	(9.75)	(12.02)	(8.14)	(4.76)	
N	28091	28091	28074	24485	24412	13055	28091
χ^2 LR	1812.3	4356.9	5214.9	3842.0	4559.2	2659.2	4161.4
Pseudo-R^2	0.047	0.114	0.136	0.115	0.137	0.147	0.109
Sample D (1960 - 1980 - 2011)							
Father's occ		2.73***	*84 cat.*	2.49***	*84 cat.*	*84 cat. + income*	3.05***
		(79.46)		(60.86)			(91.63)
Mother's occ				1.56***	*80 cat.*	*80 cat. + income*	
				(29.35)			
Grandfather's occ	2.31***	1.63***	1.50***	1.56***	1.44***	1.41***	
	(54.86)	(30.26)	(24.73)	(23.36)	(18.65)	(18.75)	
N	131194	131194	131193	98155	98140	105487	131194
χ^2 LR	3830.2	10404.2	12172.7	9130.4	10498.8	10765.3	9462.6
Pseudo-R^2	0.022	0.058	0.068	0.069	0.079	0.076	0.053
Age controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Exponentiated coefficients; t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Higher coefficients indicate higher multigenerational persistence. Son-parent-grandfather logit regressions with more detailed information on the parent generation. Outcome is whether son has white-collar occupation, ancestor's occupation is whether or not white-collar, except father and mother in columns (3) and (5) where a larger set of occupation dummies is used. One regression per column and panel; constant term and age controls not shown.

information increases the predictive power of the model as described by the χ^2 likelihood ratio tests and the pseudo-R^2.[16] There is only a moderate change in the coefficient for grandfather's occupation, which in all models remains a binary variable indicating whether the grandfather has a white-collar occupation. This provides one indication that the grandfather coefficient is not driven by imprecise measurement of the father, but we can go further.

In Model 4, we return to the binary occupational variable (Equation 6) but include a dummy variable for mother's occupation. Model 5 reports estimates using the full set of dummies (Equation 8) for both mother and father.[17] In all four time periods, the grandparent coefficient remains significant and robust to the improved measurement of parent characteristics. While there is a slight decrease in magnitude, it is small compared to the overall effect. For this reason, we conclude that the grandfather effect is indeed a reflection of latent family characteristics rather than mismeasurement of the parent's occupation.

For the third and fourth time period, we can also use data on individual incomes (from the tax records, see Online Appendix, Section B.2). The results of an estimation using individual income ranks for fathers and mothers are shown in column 6. Again, we observe that the addition of information on the parent generation does not substantially alter the estimated grandfather coefficient.

The final column reports, for reasons of comparison, results from a regression where no controls for grandfather's occupation are included. We observe that while the coefficient of father's occupation is somewhat mediated (from 14.96 to 11.79 in the first period; less in later periods) the magnitude remains similar, giving another indication that grandfather's effects do not work solely through father's observable characteristics.

In order to compare the explanatory power of the model with and without grandfathers, a starting point would be to consider the pseudo-R^2 of models (7) (fathers only) and (2) (fathers and grandfathers) in Table 6. The explanatory power of the grandfather-father model is in all cases greater than that of the father-only model. This carries over to a larger set of measures of fit typically used in evaluating logit models; for nine different ways of measuring explained variability and/or fit improvement, the model with grandfathers and fathers exhibits a higher score than the model with fathers only.[18]

A similar exercise can be performed for the other three occupation categories. These results are reported in the Online Appendix, Section A.2. In general, any significant coefficients in

[16] The increase in predictive power is substantial in absolute terms; in relative terms, the increase is only moderate, as some of the explanatory power in all models (1)-(6) is attributable to the age controls.

[17] The number of observations is lower for models 3-6 for two reasons. First, some of the detailed occupational categories have very few members in the parent generation, and observations with dummy variables that perfectly predict outcomes are dropped from the regression. Second, when including data on the mothers, we impose the same age requirements (30-60); there are also some individuals whose mother's identity is unknown. A majority of mothers in the early samples do not have a stated occupation; these are kept in the sample and assigned to an additional "homemaker" occupational category. Models with mothers also incorporate a second-degree polynomial in mother's age at time of observation. Results of regressions on balanced samples give similar results and are available on request.

[18] This is the full set of measures reported by the Stata `fitstat` command. In addition to the unadjusted McFadden R^2 reported in Table 6, these are: adjusted McFadden; McKelvey & Zavoina; Cox-Snell/ML; Cragg-Uhler/Nagelkerke; Efron; Tjur's D; and unadjusted and adjusted count R^2.

Table 3 remain significant when these more complex specifications are used, and the explanatory power of the model increases.[19] Taken together, these results suggest that the grandfather effects described in the previous sections are not primarily driven by imperfect measurement of the characteristics of the parent generation.

We now turn to a further distinction between possible mechanisms driving the influence of grandfathers, by using information on whether there was likely to have been direct contact between the generations.

5 The role of personal contact

The analysis so far has established that there is an association between grandfathers' and grandsons' social status, as measured by both occupation and income, when controlling for father's status. However, the jury is still out on whether this reflects underlying family characteristics, which even with perfect measurement would only be partly reflected in the observed status of the father, or the direct influence of the grandfather's presence during the grandson's childhood.

Zeng & Xie (2014) find that the relationship between grandparents and grandchildren's economic outcomes in rural China reflects direct interaction between generations. In their study, the effects of grandparental characteristics on grandchildren's education are strong for co-resident grandparents but nearly non-existent for non-co-resident grandparents. Does the same hold in Norway? A supplementary approach to using geographical moves to infer the direct influence of grandparents on their grandchildren is to use information on the grandparents' time of death.

This section explores the mechanisms behind the grandparental association by comparing multigenerational persistence between lineages with varying geographical distance between grandparents and grandchildren, and between lineages with different times of death of the grandfather. These factors are not completely orthogonal to the process of multigenerational transmission of occupations or income. Geographical mobility is correlated with occupational mobility, and mortality is higher in groups that are less economically advantageous. We therefore have to keep potential confounding factors in mind while performing this analysis. However, using two separate indicators of grandparental presence provides a richer picture of how multigenerational persistence mechanisms operate.

We start by comparing the grandparental persistence parameters previously reported in Table 3 in cases where individuals grew up in the same municipality as their grandparents with cases where they did not.[20] The relevant variables are the residential location of the grandfather when he was observed as an adult (time t) and the residential location of the grandson when he was observed as a child (time $t+1$). The assumption is that a geographical move sometime between

[19]If a linear probability model is used, differences in coefficient magnitudes between the reference model and the models with more details on parents' occupations is larger. However, statistically significant grandparent estimates remain so with additional controls also in the linear probability case. Results are available on request.

[20]We do not use the Zeng & Xie (2014) co-residence approach directly, as grandparental co-residence is very rare in Norway throughout the period studied here. In 1865, only seven % of households with children also had a resident grandparent; in 1910, this was down to three %.

these two observations will reduce the direct interaction between grandfather and grandson. For example, for a son in sample C who resided in Oslo municipality in 1960, we have a *non-mover dynasty* if the grandfather resided in Oslo or Aker municipalities in 1910 (the two municipalities were merged in the intervening period) and a *mover dynasty* otherwise. We then re-run the regressions from Section 3.2 on the two subsamples and compare the grandparental coefficients.

We expect less influence between generations for mover dynasties, as there is less scope for interpersonal contact. However, this loss of contact is much more likely to take place between grandfather and grandson than between father and son, as movements of young sons would likely have taken place together with the father, but not necessarily the grandfather. Hence, the coefficient on grandfather's occupation is expected to be higher (stronger persistence) for the nonmover sample than for the mover sample.

These coefficients on grandfather's occupation for the mover and nonmover subsamples are reported in Table 7, which has one column for each time period. The coefficients are obtained by interacting a dummy variable with the full set of controls in Equation (5). The first panel reports coefficients using white-collar occupation for son as outcome variable. We observe that grandparent's occupation is statistically significant for both subgroups in all time periods except for the first. The coefficient is lower for movers than for non-movers, but the difference is only of moderate size. In samples C and D, the difference between the subgroups is statistically significant. For non-movers in sample C (grandfathers in 1910, fathers in 1960, sons in 1980) we have a grandfather odds ratio (holding father's occupation constant) of 1.87. For movers, the odds ratio is 1.49. The ratio of these numbers is 1.26. Since the odds ratios correspond to exponentiated coefficients, the ratio is equivalent to the difference between the non-exponentiated coefficients, and the t-statistic of 2.6 can be interpreted as a t test, here rejecting a null hypothesis that persistence is equal across the two subgroups.[21]

We observe a slightly different picture for the other occupation groups. For manual occupations, the difference between movers and non-movers is not statistically significant. For farmers, on the other hand, there is higher persistence for non-movers than for movers. This is not surprising, as the farmer occupation is usually connected to a specific farm location. If the father had already moved away from the area of the ancestral farm when the child was born, the tie to the farming occupation is likely to have been substantially weakened.[22]

The extent to which grandfathers participate in the upbringing of their grandchildren may also depend on how far apart they live during the grandchild's formative years, even if they do not reside in the exact same municipality. An analysis of the relationship of the geographical distance between grandchild (growing up) and grandfather (measured in kilometers) shows no systematic relationship across occupations beyond that depicted in Table 7.

Next, we consider time of grandfather's death and how it relates to occupational persistence

[21] All results are qualitatively similar if we instead use the model with a full set of controls in the parent generation (see Table A16) or if a linear probability model is used (available on request).

[22] Farmers are also less likely to move. Compared to white-collar occupations, the difference in movement propensity is 27 percentage points in sample A (14 versus 41 per cent) decreasing to 18 percentage points in sample D (20 versus 38 per cent).

Table 7: Comparison of grandparental persistence for movers and non-movers

	(1)	(2)	(3)	(4)
Sample	A	B	C	D
	1865-1910	1865-1960	1910-1980	1960-2011
	Occupation: White collar			
Non-movers	2.928***	2.342***	1.875***	1.642***
	(2.768)	(4.401)	(11.961)	(25.429)
Movers	2.143	2.669***	1.487***	1.520***
	(1.575)	(4.206)	(5.523)	(14.308)
Difference	1.366	0.877	1.261***	1.080**
	(0.503)	(-0.432)	(2.600)	(2.195)
	Occupation: Farmer			
Non-movers	1.778***	1.531***	2.192***	4.185***
	(3.209)	(4.839)	(11.111)	(24.272)
Movers	0.897	1.385**	1.289*	2.199***
	(-0.293)	(2.027)	(1.947)	(5.182)
Difference	1.981*	1.105	1.701***	1.903***
	(1.663)	(0.548)	(3.584)	(3.947)
	Occupation: Manual, skilled			
Non-movers	2.229***	1.406***	0.929	0.992
	(3.413)	(2.814)	(-2.062)	(-0.527)
Movers	1.729	1.133	1.116*	1.120***
	(1.379)	(0.762)	(1.802)	(4.350)
Difference	1.289	1.241	0.833	0.886
	(0.551)	(1.061)	(-2.596)	(-4.063)
	Occupation: Manual, unskilled			
Non-movers	1.103	1.213*	1.487***	1.663***
	(0.712)	(1.842)	(5.756)	(16.594)
Movers	0.735	1.381*	1.931***	1.622***
	(-1.223)	(1.800)	(4.453)	(8.634)
Difference	1.501	0.878	0.770	1.026
	(1.415)	(-0.626)	(-1.602)	(0.398)
Son observed:	1910	1960	1980	2011
Father observed:	1900	1910	1960	1980
Grandfather observed:	1865	1865	1910	1960

Exponentiated coefficients; *t* statistics in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Note: Higher coefficients indicate higher multigenerational persistence. Table shows odds ratio coefficients on grandfather's occupation. Dependent variable: dummy for son's occupation. Dependent variables are father's and grandfather's occupation; with the regression giving a separate set of coefficient for movers (between grandfather's and grandson's observation) and non-movers. "Difference" is linear difference between parameters (log odds ratios), i.e. the ratio of the two displayed coefficients.

across generations. Digitized death records are only available from 1960 onwards, limiting this analysis to sample D (1960-1980-2011). We do a subsample analysis similar to the one in the previous subsection. The sample is split on whether the grandfather (whose occupation was recorded in 1960) is still alive in 1980, when the father's occupation is recorded. In no cases are the grandparent coefficients statistically significant, indicating no difference in persistence by whether the grandfather was alive to directly influence the grandson. The standard errors for white-collar, manual skilled and manual unskilled occupations are low, while the zero for farmers is less exact. This result holds up if the sample is instead split according to whether the grandfather survived until 2011, or if the full set of dummies for mother's and father's occupations is used.[23] Similarly, a regression using the number of years in childhood when the grandfather was alive (see Online Appendix, Section B.3) does not give any indication of a strong relationship.

Braun & Stuhler (2018) and Dribe & Helgertz (2016) have previously investigated the relationship between grandfathers' death dates and multigenerational persistence. In these cases, because of small sample sizes, the authors did not attach much weight to the absence of significant effects. However, the relatively exact zeros observed in the present case can be seen as weakening the hypothesis that social interaction with grandfathers contributes substantially to their grandchildren's occupational choice.

6 Concluding comments

The present paper has shown that grandfathers' characteristics matter for their grandsons' occupational outcomes. These results hold across a wide range of historical and institutional settings. The association is robust to several different ways of measuring economic characteristics, remains significant when we measure the status of parents in a more detailed way, and is found across subsamples split by geographical movement or grandparental mortality.

Multigenerational occupational persistence is observed not only for white-collar occupations, but also for farmers and for skilled and unskilled manual workers. The results are clear cut for white-collar occupations. Persistence is high across generations, for all generations studied spannning grandfathers born between 1805 and 1930 and grandsons born between 1870 and 1981. The relationship holds when controlling for father's occupation. There is also evidence that great-grandfathers have an influence. The magnitude of the coefficient on grandfather's occupation is up to one third of the coefficient on father's occupation. This is high, given that only one grandfather is observed; data on the maternal grandfather is not available in this study. High persistence is confirmed by the use of income rank data in the two final samples.

For farmers and for manual occupations (skilled and unskilled) the results are more nuanced. There are some periods where a significant grandfather coefficient is not obtained. Skilled workers appear to show high persistence in the early period in particular, while results for unskilled

[23]The regression coefficients are shown in the Online Appendix, Tables A17-A18.

workers are more pronounced in later samples.

There are differences across occupational groups. White-collar persistence is evident in all periods but is particularly strong in the late nineteenth century; it appears to decrease throughout the twentieth century. The high association parameters observed for white-collar occupations hint at some sort of "elite persistence". This is also reflected in analyses where a narrower definition of high-status occupations is used (the best-paid 11% or 22% of occupations in any given time period). On the other hand, unskilled and low-status occupations show a tendency towards increasing persistence over time.

Farmers always exhibit strong persistence. While this is not surprising, it should be kept in mind in any study of historical intergenerational mobility. In most countries outside the most industrialized part of Western Europe, a substantial proportion of the population were farmers well into the twentieth century, and studies of mobility that rely on imputed status or income for this period are likely to be strongly affected by trends in mobility into and out of farming.

Families staying in the same place across generations show slightly higher multigenerational persistence. However, there is substantial persistence also in dynasties that move between observations. The difference allows for some effect due to direct personal interaction between grandfather and grandson but may also reflect shared personal networks or region-specific competencies. The timing of grandfather's death does not influence the degree of multigenerational persistence, suggesting that the role of personal interaction is limited.

This paper documents that there are substantial differences in the strength of persistence over time, from the nineteenth through to the twenty-first century. During this time period, Norway grew wealthier and education, health and social insurance policies were substantially expanded. At the same time, persistence in white-collar and manual skilled occupations fell substantially. Hence, while persistence is found in all time periods, there still appears to be a role for institutions in the transmission of social status across generations, as we observe substantial changes in the level of measured persistence over time. Moreover, the differences across occupational groups suggest that both vertical (status) and horizontal (sector; notably agricultural/non-agricultural) must be taken into account. These explanations do not rule out unobservable inherited characteristics as an important channel of influence. The changes over time do however not support the argument by Clark (2014) that such latent characteristics are strong and virtually unchanged over time, regardless of the institutional framework.

Non-manifested genetic traits (inheritance) may explain some of the dynasty persistence. However, traits may well be manifested in the father and yet not be reflected in his choice of occupation. In most cases, people cannot find a job in which all their skills are useful, and the son of a carpenter may well find joy in carpentry (and pass this on to his son through social interaction) even though he ends up working in another occupation. The social network of the family may also reflect the ancestors' economic life and affect the occupational choice and success of the grandson. For some occupations, even outside farming, direct inheritance and family wealth may play a role as well. What we certainly learn is that persistence in economic outcomes spans more than two generations.

References

Aaberge, Rolf, Atkinson, Anthony B., & Modalsli, Jørgen. 2020. Estimating long-run income inequality from mixed tabular data: Empirical evidence from Norway, 1875-2017. *Journal of Public Economics*, **187**, 104–196.

Abramitzky, Ran, Boustan, Leah Platt, Eriksson, Katherine, Feigenbaum, James J, & Perez, Santiago. 2020. Automated Linking of Historical Data. *NBER working paper*, **25825**.

Agresti, Alan. 2002. *Categorical Data Analyis*. Wiley Interscience.

Azam, Mehtabul. 2015. Intergenerational Occupational Mobility among Men in India. *Journal of Development Studies*, **51**(10).

Bailey, Martha, Cole, Connor, Henderson, Morgan, & Massey, Catherine. 2019. How well do automated linking methods perform? Lessons from U.S. historical data. *NBER working paper*, **24019**.

Bailey, Martha, Cole, Connor, & Massey, Catherine. 2020. Simple strategies for improving inference with linked data: a case study of the 1850-1930 IPUMS linked representative historical samples. *Historical Methods*, **53**(2), 80–93.

Boberg-Fazlic, Nina, & Sharp, Paul. 2018. North and south: long-run social mobility in England and attitudes toward welfare. *Cliometrica*, **12**(2), 251–276.

Bolt, Jutta, & van Zanden, Jan Luiten. 2013. The First Update of the Maddison Project: Re-Estimating Growth Before 1820. *Maddison-Project Working Paper*, **4**.

Bratberg, Espen, Nilsen, Øivind Anti, & Vaage, Kjell. 2005. Intergenerational earnings mobility in Norway: Levels and trends. *Scandinavian Journal of Economics*, **107**(3), 419–435.

Braun, Sebastian Till, & Stuhler, Jan. 2018. The Transmission of Inequality Across Multiple Generations: Testing Recent Theories with Evidence from Germany. *Economic Journal*, **128**, 576–611.

Buis, Maarten L. 2016. Logistic regression: When can we do what we think we can do? *Mimeo*.

Chan, T. W., & Boliver, V. 2013. The Grandparents Effect in Social Mobility: Evidence from British Birth Cohort Studies. *American Sociological Review*, **78**(4), 662–678.

Clark, Gregory. 2014. *The Son Also Rises: Surnames and the History of Social Mobility*. Princeton University Press.

Clark, Gregory, & Cummins, Neil. 2015. Intergenerational wealth mobility in England, 1858-2012: Surnames and social mobility. *Economic Journal*, **125**(582), 61–85.

Dribe, Martin, & Helgertz, Jonas. 2016. The lasting impact of grandfathers: Class, occupational status, and earnings over three generations in Sweden 1815-2011. *Journal of Economic History*, **76**(4), 969–1000.

Ferrie, Joseph, Massey, Catherine, & Rothbaum, Jonathan. 2016. Do grandparents and great-grandparents matter? Multigenerational mobility in the US, 1910-2013. *NBER Working Paper 22635*.

Ferrie, Joseph, Massey, Catherine, & Rothbaum, Jonathan. 2020. Do Grandparents Matter? Multigenerational mobility in the U.S., 1940-2015. *mimeo*.

Guell, M., Rodriguez Mora, J. V., & Telmer, C. I. 2015. The Informational Content of Surnames, the Evolution of Intergenerational Mobility, and Assortative Mating. *The Review of Economic Studies*, **82**(2), 693–735.

Hertel, Florian R, & Groh-Samberg, Olaf. 2014. Class mobility across three generations in the U.S. and Germany. *Research in Social Stratification and Mobility*, **35**, 35–52.

Hodge, Robert W. 1966. Occupational mobility as a probability process. *Demography*, **3**(1), 19–34.

Hout, Michael, & Guest, Avery. 2013. Intergenerational Occupational Mobility in Great Britain and the United States since 1850: Comment. *American Economic Review*, **103**(5), 2021–2040.

Knigge, Antonie. 2016. Beyond the Parental Generation: The Influence of Grandfathers and Great-grandfathers on Status Attainment. *Demography*, **53**, 1219–1244.

Kroeger, Sarah, & Thompson, Owen. 2015. Educational Mobility Across Three Generations of American Women. *Economics of Education Review*, **53**, 72–86.

Lindahl, Mikael, Palme, Mårten, Massih, Sofia Sandgren, & Sjögren, Anna. 2015. Long-term intergenerational persistence of human capital: an empirical analysis of four generations. *Journal of Human Resources*, **50**(1), 1–33.

Long, Jason, & Ferrie, Joseph. 2013a. Intergenerational Occupational Mobility in Great Britain and the United States since 1850. *American Economic Review*, **103**(4), 1109–1137.

Long, Jason, & Ferrie, Joseph. 2013b. Intergenerational Occupational Mobility in Great Britain and the United States since 1850: Reply. *American Economic Review*, **103**(5), 2041–2049.

Long, Jason, & Ferrie, Joseph. 2018. Grandfathers Matter (ed): Occupational Mobility Across Three Generations in the US and Britain, 1850-1911. *Economic Journal*, **128**(612), F422–F445.

Mare, Robert D. 2011. A multigenerational view of inequality. *Demography*, **48**(1), 1–23.

Modalsli, Jørgen. 2017. Intergenerational mobility in Norway, 1865-2011. *Scandinavian Journal of Economics*, **119**(1), 34–71.

Mood, Carina. 2010. Logistic regression: Why we cannot do what we think we can do, and what we can do about it. *European Sociological Review*, **26**(1), 67–82.

Nilsen, Øivind Anti, Vaage, Kjell, Aakvik, Arild, & Jacobsen, Karl. 2012. Intergenerational earnings mobility revisited: Estimates based on lifetime earnings. *Scandinavian Journal of Economics*, **114**(1), 1–23.

Norwegian Department of Justice. 1871. *Tabeller til oplysning om stemmerets- indtægts- og skatteforholdene i Norge i aaret 1868.* Norges Officielle Statistik B no. 3.

Olivetti, Claudia, Paserman, Daniele M., & Salisbury, Laura. 2018. Three-generation mobility in the United States, 1850-1940: The role of maternal and paternal grandparents. *Explorations in Economic History*, **70**, 73–90.

Pekkarinen, Tuomas, Salvanes, Kjell G, & Sarvimaki, Matti. 2017. The evolution of social mobility: Norway over the 20th century. *Scandinavian Journal of Economics*, **119**(1), 5–33.

Semmingsen, Ingrid. 1954. Standssamfunnets oppløsning i Norge. *Ståndssamhällets upplösning i Norden*, **1**(2), 49–86.

Solon, Gary. 2018. What do we know so far about multigenerational mobility? *Economic Journal*, **128**(612), F340–F352.

Statistics Norway. 1915. *Indtægts- og formuesforhold efter skatteligningen 1911 i forbindelse med Folketællingen 1910.* Norges Officielle Statistik VI no. 24.

Warren, John Robert, & Hauser, Robert M. 1997. Social stratification across three generations: New evidence from the Wisconsin Longitudinal Study. *American Sociological Review*, **62**(4), 561–572.

Xie, Yu, & Killewald, Alexandra. 2013. Intergenerational Occupational Mobility in Great Britain and the United States Since 1850: Comment. *American Economic Review*, **103**(5), 2003–2020.

Zeng, Zhen, & Xie, Yu. 2014. The Effects of Grandparents on Children's Schooling: Evidence From Rural China. *Demography*, **51**(2), 599–617.

Online Appendix to "Multigenerational persistence: Evidence from 146 years of administrative data"

Jørgen Modalsli

February 28, 2021

This Online Appendix must be read together with the main paper (*Multigenerational persistence: Evidence from 146 years of administrative data*).

Section A gives supplemental information on the data used and the construction of links, and provides extensions (supplemental tables) to the analysis in the main text.

Section B goes into detail on three more extensive ways of analysing the data: First, how incorporating information outside the matrix diagonal does not substantially alter the baseline results; second, how persistence is found also when data on income is used (rather than occupation), and third, that grandparental presence also has an impact when measured on the intensive (rather than extensive) margin.

A Data, matching and robustness analyses

A.1 Data overview

The data are explained in Sections 2.1-2.2 of the main paper.

Data linkage: Before 1960, records are linked using information on name, birth dates and birth times. Care is taken only to link on the basis of information that is time-invariant; i.e., information on residence, marital status etc. is not used.

As last names were not fixed by law in Norway until the 1920s, the last names given in the censuses of 1865, 1900 and 1910 are supplemented by constructed patronymics (based on the first name of the co-resident father) and place names when attempting to match names, as the reported names could change over time for the same individual. Similarities for first and last names are evaluated using the Levensthein algorithm, giving a score for name similarity. For the last name, the best alternative of the given last name, constructed patronymic and place name is used.

Municipality of birth is given in the 1865, 1900 and 1910 censuses. The highest score is given to candidate matches where the municipalities match; however, candidates reported as having been born in the same county are also considered. In 1960, only the county of birth combined with whether a person was born in an urban or rural locality is stated. Similar information is constructed from the 1910 census and used in the 1910-1960 linkage.

The censuses of 1865 and 1900 only state year of birth. The 1910 and 1960 censuses contain full birth dates and this information is used in the 1910-1960 link. In all cases, some discrepancy is allowed but gives candidate matches a lower score.

An aggregate score is constructed for all candidate matches (in principle, all pairs of observations for two censuses) and considered acceptable if the score is high (i.e. most information matches) and the match is unique (i.e. there are no other good matches for either candidate). The matching algorithm is explained in more detail in the appendix to Modalsli (2017).

Transition matrices: See Table A1 for the three-generation transition matrices.

Relationship between administrative and survey data: With the exception of some of the studies referred to in the Introduction, research on society-wide, long-run multigenerational persistence is typically based on survey data. In most cases, the middle generation is interviewed and asked about the characteristics of their parent(s) when they were growing up. Then data on the child are either reported by the parent directly (when only one interview is conducted at a time when the child is old enough to have entered the labor market) or collected later in follow-up rounds. While surveys often collect information that is not available in administrative data and may incorporate retrospective information about events prior to the interviews, there are challenges implicit in how individuals remember past events and how recall error and non-response are distributed across social groups. Blau & Duncan (1967, Appendix D-F) discuss in detail the extent to which retrospective responses in the 1962 OCG survey study (which only covers two generations) are consistent with administrative data available from the U.S. Census. They conclude that there is likely to be some response bias in survey data. In the OCG's Chicago Pretest Matching Study, of a subsample of 570 individuals, 485 completed the questionnaire, and 342 reported the place they lived during childhood. Of the 137 of this last group the research team was able to find in the census records, there was a discrepancy with respect to father's occupation of 30% (with detailed occupational groupings) and 8% (with four groups). Some discrepancy is also found when 1920-1940 male occupational distributions, estimated from fathers' occupations reported in the OCG, are compared with the actual census distribution. While some of these results may be due to the short-term occupational mobility of fathers, Blau and Duncan conclude (p. 469) that "although some of the difference is a result of upward mobility of fathers, some of it probably does reflect response bias".

Mayer (2007) reviews the literature on retrospective questions and concludes that while the quality of the survey process does affect the degree of recall error, this error can probably not be completely eliminated. For this reason, administrative data should be used to verify any results that are obtained using survey methodology. Moreover, few surveys exist before the 1950s, and even today, sample sizes are often small.

Grandfather's occupation	Father's occupation	Son's occupation	Sample A	Sample B	Sample C	Sample D
White collar	White collar	White collar	41	150	2,125	15,682
		Farmer	1	5	28	56
		Manual, skilled	8	26	395	2,956
		Manual, unskilled	3	8	50	940
	Farmer	White collar	1	9	98	193
		Farmer	1	12	93	53
		Manual, skilled	3	10	90	93
		Manual, unskilled	3	2	20	22
	Manual, skilled	White collar	3	21	433	3,703
		Farmer	0	2	8	33
		Manual, skilled	9	25	380	2,044
		Manual, unskilled	1	0	41	588
	Manual, unskilled	White collar	0	3	73	301
		Farmer	1	0	7	5
		Manual, skilled	3	11	53	164
		Manual, unskilled	0	5	18	64
Farmer	White collar	White collar	38	202	1,195	5,531
		Farmer	8	25	61	160
		Manual, skilled	27	87	399	1,898
		Manual, unskilled	7	21	47	481
	Farmer	White collar	67	392	1,798	3,354
		Farmer	347	1,209	2,141	1,350
		Manual, skilled	157	736	2,676	2,718
		Manual, unskilled	575	368	385	639
	Manual, skilled	White collar	22	142	1,285	5,591
		Farmer	14	35	87	359
		Manual, skilled	58	248	1,889	5,538
		Manual, unskilled	16	22	145	1,309
	Manual, unskilled	White collar	9	43	411	607
		Farmer	17	34	53	102
		Manual, skilled	33	114	726	557
		Manual, unskilled	61	79	333	222
Manual, skilled	White collar	White collar	25	90	1,340	16,862
		Farmer	0	3	13	76
		Manual, skilled	12	34	468	5,232
		Manual, unskilled	2	5	34	1,510
	Farmer	White collar	2	10	140	313
		Farmer	4	36	110	68
		Manual, skilled	4	29	251	285
		Manual, unskilled	9	13	29	66
	Manual, skilled	White collar	10	70	1,665	16,076
		Farmer	1	6	23	173
		Manual, skilled	49	169	2,172	12,887
		Manual, unskilled	11	11	160	3,244
	Manual, unskilled	White collar	2	13	129	996
		Farmer	0	7	2	12
		Manual, skilled	6	45	177	759
		Manual, unskilled	4	3	34	283
Manual, unskilled	White collar	White collar	7	56	302	2,858
		Farmer	3	5	4	16
		Manual, skilled	9	31	127	984
		Manual, unskilled	2	7	19	364
	Farmer	White collar	7	70	148	171
		Farmer	38	209	130	55
		Manual, skilled	22	221	333	167
		Manual, unskilled	82	102	66	50
	Manual, skilled	White collar	18	76	571	3,432
		Farmer	10	18	13	88
		Manual, skilled	59	247	1,004	3,063
		Manual, unskilled	18	38	85	1,014
	Manual, unskilled	White collar	11	33	217	1,125
		Farmer	12	55	24	44
		Manual, skilled	37	195	480	874
		Manual, unskilled	76	87	278	734
Total			2,086	6,040	28,091	131,194

Table A1: Three-generation occupational transitions: Number of individuals tabulated by grandfather's, father's and son's occupation, four samples (see Table 1 for sample definitions)

A.2 Regression results, grandfather-father-son regressions

This Appendix section shows some additional tables referred to in the main text of the paper.

Detailed occupational regressions: Table A2 shows results for more specific occupational categories. It is evident from the table that in most cases, using smaller occupational groups yields coefficients of comparable magnitude and with statistical significance. Some difference from the baseline specification is expected.

Models with controls: Tables A3-A5 report the same results as Table 6 for farmers, skilled manual occupations and unskilled manual occupations.

	(1)	(2)	(3)	(4)
Sample	A	B	C	D
	1865-1910	1865-1960	1910-1980	1960-2011
Occupation: Doctors (subset of White collar)				
Father			30.43***	17.36***
			(19.27)	(31.55)
Grandfather		23.01***	4.170***	2.350***
		(2.83)	(3.84)	(5.14)
Occupation: Salespeople (subset of White collar)				
Father	89.64***		1.894	2.106***
	(3.30)		(1.40)	(7.74)
Grandfather				1.631***
				(3.34)
Occupation: Carpenters (subset of Manual, skilled)				
Father	7.406***	2.220**	4.306***	3.563***
	(4.16)	(2.46)	(14.15)	(26.02)
Grandfather		2.257**	1.136	1.416***
		(2.01)	(0.71)	(5.74)
Occupation: Caretakers (subset of Manual, unskilled)				
Father	747.1***		3.330**	1.349
	(3.38)		(2.35)	(1.10)
Grandfather				1.042
				(0.12)
Occupation: Fishermen (subset of Manual, unskilled)				
Father	10.31***	14.33***	46.87***	12.66***
	(7.15)	(12.91)	(39.68)	(27.45)
Grandfather	2.153	2.035**	1.773***	7.291***
	(1.48)	(2.23)	(4.62)	(21.87)
Age controls	Yes	Yes	Yes	Yes
N	2086	6040	28091	131194
Son observed:	1910	1960	1980	2011
Father observed:	1900	1910	1960	1980
Grandfather observed:	1865	1865	1910	1960

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A2: Odds ratio coefficients for binary occupational regressions on Father's and grandfather's occupations. Separate regressions for each sample and occupational category. Constant terms and coefficients on quadratic controls for age for all three generations were also included in the regressions. Blank cells denote occupation-sample combinations where there were too few observations to estimate coefficients (i.e., not all 8 combinations in sufficient detail).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sample A (1865 - 1900 - 1910)							
Father's occ		3.69***	*41 cat.*	3.99***	*41 cat.*		4.48***
		(8.31)		(7.16)			(10.39)
Mother's occ				(omitted)	*11 cat.*		
				(.)			
Grandfather's occ	2.86***	1.60***	1.84***	1.46**	1.67***		
	(7.41)	(2.96)	(3.88)	(1.97)	(2.65)		
N	2086	2086	1952	1270	1184	–	2086
χ^2 LR	147.9	228.0	207.1	157.2	157.2		218.9
Pseudo-R^2	0.067	0.104	0.097	0.112	0.116		0.100
Sample B (1865 - 1910 - 1960)							
Father's occ		8.18***	*63 cat.*	7.94***	*63 cat.*		9.36***
		(24.49)		(23.22)			(27.18)
Mother's occ				0.66	*22 cat.*		
				(-0.96)			
Grandfather's occ	2.90***	1.47***	1.61***	1.44***	1.58***		
	(15.77)	(5.10)	(6.36)	(4.64)	(5.86)		
N	6040	6040	5763	5574	5167	–	6040
χ^2 LR	284.1	1076.6	1027.4	970.5	839.8		1050.2
Pseudo-R^2	0.040	0.152	0.148	0.148	0.133		0.148
Sample C (1910 - 1960 - 1980)							
Father's occ		18.71***	*67 cat.*	20.47***	*67 cat.*	*67 cat. + income*	24.63***
		(44.78)		(41.47)			(52.12)
Mother's occ				5.39	*47 cat.*	*47 cat. + income*	
				(1.60)			
Grandfather's occ	6.44***	1.93***	1.87***	1.91***	1.85***	1.94***	
	(34.70)	(10.84)	(10.23)	(9.82)	(9.26)	(6.89)	
N	28091	28091	27300	24485	23223	11799	28091
χ^2 LR	1977.6	5069.3	5098.1	4500.3	4418.1	2096.3	4944.9
Pseudo-R^2	0.109	0.278	0.283	0.283	0.283	0.283	0.272
Sample D (1960 - 1980 - 2011)							
Father's occ		8.64***	*84 cat.*	8.06***	*84 cat.*	*84 cat. + income*	18.30***
		(43.94)		(36.82)			(68.23)
Mother's occ				1.45***	*80 cat.*	*80 cat. + income*	
				(5.52)			
Grandfather's occ	9.41***	3.92***	3.11***	3.93***	3.05***	3.20***	
	(48.73)	(25.13)	(20.53)	(22.45)	(17.85)	(17.67)	
N	131194	131194	129246	98155	95186	103482	131194
χ^2 LR	3517.5	5595.8	6001.8	4787.2	5140.9	4033.3	4953.1
Pseudo-R^2	0.136	0.216	0.232	0.227	0.245	0.222	0.191
Age controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Exponentiated coefficients; t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A3: Son-parent-grandfather logit regressions with more detailed information on the parent generation (cf. Table 6), Farmers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sample A (1865 - 1900 - 1910)							
Father's occ		5.46***	*41 cat.*	5.46***	*41 cat.*		6.18***
		(12.26)		(9.05)			(13.59)
Mother's occ				5.80**	*11 cat.*		
				(2.12)			
Grandfather's occ	3.64***	2.04***	1.79***	2.48***	2.29***		
	(7.18)	(3.55)	(2.63)	(3.49)	(2.65)		
N	2086	2086	2065	1271	1232	–	2086
χ^2 LR	82.3	232.0	327.1	149.8	189.1		219.7
Pseudo-R^2	0.036	0.101	0.144	0.113	0.149		0.096
Sample B (1865 - 1910 - 1960)							
Father's occ		3.31***	*63 cat.*	3.15***	*63 cat.*		3.46***
		(17.07)		(15.56)			(18.11)
Mother's occ				0.45	*22 cat.*		
				(-1.89)			
Grandfather's occ	1.88***	1.32***	1.29**	1.33***	1.28**		
	(6.94)	(2.84)	(2.42)	(2.81)	(2.20)		
N	6040	6040	6005	5574	5521	–	6040
χ^2 LR	55.3	354.6	580.5	315.2	523.7		346.6
Pseudo-R^2	0.007	0.045	0.073	0.043	0.072		0.044
Sample C (1910 - 1960 - 1980)							
Father's occ		2.35***	*67 cat.*	2.33***	*67 cat.*	*67 cat. + income*	2.33***
		(31.97)		(29.35)			(32.93)
Mother's occ				1.07	*47 cat.*	*47 cat. + income*	
				(0.63)			
Grandfather's occ	1.25***	0.96	1.14***	0.94	1.11***	1.14***	
	(7.93)	(-1.38)	(3.90)	(-1.79)	(2.95)	(2.62)	
N	28091	28091	28074	24485	24398	13055	28091
χ^2 LR	149.8	1188.5	2563.1	1016.6	2248.9	1459.3	1186.6
Pseudo-R^2	0.004	0.031	0.067	0.031	0.068	0.084	0.031
Sample D (1960 - 1980 - 2011)							
Father's occ		2.17***	*84 cat.*	2.13***	*84 cat.*	*84 cat. + income*	2.18***
		(62.26)		(51.77)			(63.43)
Mother's occ				1.12***	*80 cat.*	*80 cat. + income*	
				(3.95)			
Grandfather's occ	1.16***	1.02*	1.09***	1.02	1.09***	1.09***	
	(12.37)	(1.68)	(6.61)	(1.42)	(5.78)	(6.17)	
N	131194	131194	131182	98155	98130	105479	131194
χ^2 LR	409.1	4351.7	7028.8	3485.2	6159.1	6578.7	4348.9
Pseudo-R^2	0.003	0.027	0.043	0.029	0.051	0.051	0.027
Age controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Exponentiated coefficients; t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A4: Son-parent-grandfather logit regressions with more detailed information on the parent generation (cf. Table 6), Manual, skilled

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sample A (1865 - 1900 - 1910)							
Father's occ		1.69***	*41 cat.*	1.98***	*41 cat.*		1.65***
		(3.80)		(3.81)			(3.80)
Mother's occ				1.05	*11 cat.*		
				(0.27)			
Grandfather's occ	1.07	0.93	1.33**	0.89	1.21		
	(0.58)	(-0.57)	(2.20)	(-0.76)	(1.15)		
N	2086	2086	2016	1271	1209	–	2086
χ^2 LR	15.1	29.6	271.0	29.1	169.7		29.3
Pseudo-R^2	0.005	0.010	0.098	0.017	0.103		0.010
Sample B (1865 - 1910 - 1960)							
Father's occ		2.33***	*63 cat.*	2.23***	*63 cat.*		2.48***
		(8.32)		(7.52)			(9.29)
Mother's occ				1.21	*22 cat.*		
				(1.35)			
Grandfather's occ	1.46***	1.22**	1.43***	1.26**	1.45***		
	(4.44)	(2.25)	(3.98)	(2.48)	(3.97)		
N	6040	6040	5809	5574	5285	–	6040
χ^2 LR	41.5	105.0	215.1	101.3	193.7		100.1
Pseudo-R^2	0.009	0.023	0.047	0.024	0.046		0.022
Sample C (1910 - 1960 - 1980)							
Father's occ		5.58***	*67 cat.*	5.62***	*67 cat.*	*67 cat. + income*	6.16***
		(30.68)		(28.67)			(33.60)
Mother's occ				1.12	*47 cat.*	*47 cat. + income*	
				(0.54)			
Grandfather's occ	2.33***	1.55***	1.52***	1.62***	1.58***	1.39***	
	(14.54)	(7.06)	(6.48)	(7.19)	(6.54)	(3.30)	
N	28091	28091	27897	24485	24095	12906	28091
χ^2 LR	300.1	1120.6	1484.1	1035.0	1369.2	645.9	1073.4
Pseudo-R^2	0.023	0.086	0.114	0.091	0.121	0.114	0.082
Sample D (1960 - 1980 - 2011)							
Father's occ		2.21***	*84 cat.*	2.19***	*84 cat.*	*84 cat. + income*	2.60***
		(23.37)		(19.78)			(29.32)
Mother's occ				1.18***	*80 cat.*	*80 cat. + income*	
				(5.95)			
Grandfather's occ	1.91***	1.65***	1.33***	1.70***	1.35***	1.28***	
	(25.14)	(18.73)	(9.97)	(16.96)	(8.97)	(7.70)	
N	131194	131194	131187	98155	98072	105432	131194
χ^2 LR	646.6	1129.3	2471.3	1019.5	2201.1	2106.4	804.3
Pseudo-R^2	0.008	0.014	0.032	0.018	0.038	0.034	0.010
Age controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Exponentiated coefficients; t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A5: Son-parent-grandfather logit regressions with more detailed information on the parent generation (cf. Table 6), Manual, unskilled

A.3 Great-grandfathers and sample selection

See Section 3.4 of the main paper. Table A6 shows how sample selection affects the results shown in Table 5.

	(1)	(2)	(3)	(4)	(5)	(6)
Sample	C 1865-1980	C 1865-1980	D 1910-2011	D 1910-2011	D 1865-2011	D 1865-2011
	with data on 4 gen.		with data on 4 gen.		with data on 5 gen.	
White collar						
Father	5.681***	5.566***	2.917***	2.894***	3.092***	3.087***
	(14.28)	(14.04)	(31.83)	(31.51)	(9.67)	(9.64)
Grandfather	1.792***	1.622***	1.637***	1.559***	1.747***	1.751***
	(3.62)	(2.83)	(11.78)	(10.05)	(3.40)	(3.40)
Great-grandfather		1.532*		1.185***	0.994	0.916
		(1.79)		(3.23)	(-0.03)	(-0.40)
Great-great-grandfather						1.414
						(1.03)
Farmer						
Father	23.57***	23.15***	8.280***	8.049***	6.963***	6.677***
	(13.16)	(13.05)	(20.29)	(20.07)	(6.46)	(6.33)
Grandfather	1.575**	1.426	4.542***	3.873***	2.629**	2.655**
	(2.17)	(1.62)	(11.35)	(9.69)	(2.51)	(2.54)
Great-grandfather		1.337		1.525***	1.512	1.213
		(1.62)		(3.36)	(1.04)	(0.47)
Great-great-grandfather						1.812
						(1.64)
Manual, skilled						
Father	2.753***	2.756***	2.224***	2.222***	2.166***	2.159***
	(10.72)	(10.71)	(24.65)	(24.60)	(7.01)	(6.97)
Grandfather	0.800*	0.813	1.025	1.078**	1.194	1.198
	(-1.75)	(-1.59)	(0.71)	(2.13)	(1.45)	(1.47)
Great-grandfather		0.897		0.810***	0.872	0.898
		(-0.61)		(-5.13)	(-0.76)	(-0.58)
Great-great-grandfather						0.810
						(-0.86)
Manual, unskilled						
Father	4.812***	4.776***	2.165***	2.138***	0.968	0.944
	(7.86)	(7.80)	(8.71)	(8.53)	(-0.08)	(-0.14)
Grandfather	1.417	1.291	1.520***	1.487***	1.150	1.151
	(1.48)	(1.06)	(5.64)	(5.21)	(0.47)	(0.47)
Great-grandfather		1.371		1.113	1.727**	1.584*
		(1.63)		(1.45)	(2.10)	(1.72)
Great-great-grandfather						1.320
						(1.32)
Age controls	Yes	Yes	Yes	Yes	Yes	Yes
N	2422	2422	19700	19700	1676	1676

Exponentiated coefficients; t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A6: Four and five generations, samples C and D. This table compares the results in Table 5 with one-generation-less estimates for the same restricted samples.

A.4 Weighting observations by match propensity

This section presents the results for the sample weighted by match probability.

As explained in the main text, in the initial stage the unit of observation is a father-son pair and an indicator variable $M \in (0, 1)$ is constructed, denoting a successful match to a grandfather. This is then as outcome in a logit regression where the right-hand side variables are the father's occupation category, father's age and log of the population size of father's municipality. From the estimated relationship a match probability can be predicted for the relevant population, and the inverse of these probabilities are then used as weights in the matched sample.

In principle, this could also be done in the other direction (e.g. using the grandfather generation as base, and see if they are linked to the father generation. This would, however, also incorporate differences due to differential fertility. While this is a topic that could deserve further study, it is beyond the scope of the present paper.

Table A7 compares the weighted and unweighted regressions. The coefficients are in general similar, though some differences stand out (primarily the weighted coefficients in sample B). Some of the cell sizes are very small, giving undue weight to a very small set of individuals from some occupation categories (most notably farmers). For this reason, one would not want to weight in this way in the baseline specification. The results show that the overall trends discussed in this paper are not artifacts of the matching process, though the preciseness of sample B, where the measurement of the first two generations are only 10 years apart, appears to be less robust to changing specifications than the other samples.

Sample	A	A	B	B	C	C	D	D
	Unweight.	Weighted	Unweight.	Weighted	Unweight.	Weighted	Unweight.	Weighted
Occupation: White collar								
Father	11.79***	13.13***	8.071***	36.38***	5.151***	5.004***	2.730***	2.696***
	(13.58)	(13.79)	(23.28)	(7.71)	(48.16)	(46.61)	(79.46)	(51.53)
Grandf.	2.838***	2.882***	2.504***	1.755	1.802***	1.782***	1.631***	1.657***
	(3.51)	(3.89)	(6.19)	(0.79)	(13.93)	(13.46)	(30.26)	(20.64)
Occupation: Farmers								
Father	3.686***	3.571***	8.179***	45.38***	18.71***	19.00***	8.636***	7.633***
	(8.31)	(7.54)	(24.49)	(8.71)	(44.78)	(42.79)	(43.94)	(24.73)
Grandf.	1.595***	1.723***	1.471***	2.662**	1.929***	1.938***	3.916***	4.350***
	(2.96)	(3.09)	(5.10)	(2.25)	(10.84)	(10.46)	(25.13)	(15.77)
Occupation: Manual, skilled								
Father	5.458***	4.953***	3.312***	20.37***	2.351***	2.272***	2.171***	2.137***
	(12.26)	(11.05)	(17.07)	(5.30)	(31.97)	(30.30)	(62.26)	(39.46)
Grandf.	2.039***	1.998***	1.316***	2.014	0.959	0.966	1.021*	1.019
	(3.55)	(3.26)	(2.84)	(1.20)	(-1.38)	(-1.11)	(1.68)	(0.96)
Occupation: Manual, unskilled								
Father	1.688***	1.959***	2.327***	2.576	5.581***	5.523***	2.210***	2.071***
	(3.80)	(4.63)	(8.32)	(1.42)	(30.68)	(30.33)	(23.37)	(14.18)
Grandf.	0.935	0.950	1.223**	0.556	1.554***	1.555***	1.652***	1.695***
	(-0.57)	(-0.41)	(2.25)	(-0.81)	(7.06)	(7.05)	(18.73)	(12.42)
Age cont.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	2086	2086	6040	6040	28091	28091	131194	131194

Exponentiated coefficients; t statistics in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Table A7: Coefficient from the son-father-grandfather regression for the unweighted sample (baseline) and for observations weighted by the inverse probability of matching the father-son pair to a grandfather observation.

A.5 Age at which occupation is measured

Throughout the paper, occupation is measured between ages 30 and 60. As shown in the descriptives (Table 1), not all samples have a full 30-year range of birth years for all generations. However, for the final generation, we have both a reasonable range of birth years and a large enough sample to make subsample analysis of multigenerational persistence for different age groups viable.

In this robustness check, we split the sample according to whether each generation is aged above or below 40. This gives us two subsamples of respectively 12,478 individuals where both son, father and grandfather are below 40 when occupation is measured, and 22,242 where all three are 40 or older.

The regressions from the baseline regressions are then run on both sub-samples. The results, together with the baseline, are shown in Table A8 below. It is evident that the results are robust to different timing of occupation measurement, as the coefficients are in most cases not substantially different form baseline. However, for farmers, the coefficient on grandfathers is higher.

	Occupation: White collar			Occupation: Farmer		
	Full sample	Age < 40	Age ≥ 40	Full sample	Age < 40	Age ≥ 40
Father	2.730***	2.446***	2.830***	8.636***	8.661***	9.475***
	(79.46)	(22.69)	(32.38)	(43.94)	(8.72)	(25.35)
Grandfather	1.631***	1.548***	1.740***	3.916***	5.508***	3.339***
	(30.26)	(8.58)	(12.84)	(25.13)	(7.32)	(11.94)
Age controls	Yes	Yes	Yes	Yes	Yes	Yes
N	131194	12478	22242	131194	12478	22242
	Occupation: Manual, skilled			Occupation: Manual, unskilled		
	Full sample	Age < 40	Age ≥ 40	Full sample	Age < 40	Age ≥ 40
Father	2.171***	2.113***	1.934***	2.210***	2.693***	2.250***
	(62.26)	(18.60)	(22.08)	(23.37)	(9.13)	(11.28)
Grandfather	1.021*	1.125***	0.974	1.652***	1.669***	1.608***
	(1.68)	(2.90)	(-0.86)	(18.73)	(5.99)	(7.73)
Age controls	Yes	Yes	Yes	Yes	Yes	Yes
N	131194	12478	22242	131194	12478	22242

Exponentiated coefficients; t statistics in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Table A8: Coefficient for the son-father-grandfather regression in Sample D (sons born 1960-1981), for baseline and for sample split by age of all three generations.

A.6 Persistence in high- and low-ranked occupations

See Table A9.

	(1)	(2)	(3)	(4)
Sample	A	B	C	D
	1865-1910	1865-1960	1910-1980	1960-2011
Occupations with highest income ranking (top 10-12 %)				
Father	7.088***	4.897***	3.929***	2.027***
	(9.34)	(14.84)	(30.87)	(31.32)
Grandfather	2.397***	2.145***	1.904***	1.527***
	(3.51)	(6.06)	(12.76)	(16.38)
Occupations with highest income ranking (top 22-24 %)				
Father	8.608***	4.535***	3.942***	1.945***
	(13.82)	(19.01)	(39.34)	(40.60)
Grandfather	2.468***	1.673***	1.651***	1.449***
	(4.74)	(5.32)	(14.16)	(20.78)
Occupations with lowest income ranking (bottom 22-23 %)				
Father	12.24***	1.673***	4.537***	1.902***
	(20.06)	(7.84)	(51.13)	(38.24)
Grandfather	0.794	0.780***	1.100**	1.144***
	(-1.43)	(-3.41)	(2.78)	(8.34)
Son observed:	1910	1960	1980	2011
Father observed:	1900	1910	1960	1980
Grandfather observed:	1865	1865	1910	1960

Exponentiated coefficients; t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A9: Odds ratio coefficients for binary occupational regressions on father and grandfather's occupations. Occupation categories of constant size across time periods. Dependent variable: son's occupation. Separate logit regressions for each sample and occupational category. Constant terms and coefficients on quadratic controls for age for all three generations were also included in the regressions.

B Supplemental analyses

B.1 Incorporating information from outside the matrix diagonal

In addition to the influence on son's occupation due to fathers and grandfathers having the same occupation, there may be cross-occupational effects; for example, the probability of a son entering a white-collar occupation may differ, depending on whether the father had a manual skilled or manual unskilled occupation. With four occupational categories, there are six relevant comparisons of occupations; we restrict the analysis to similar comparisons for father and grandfather, giving a total of 36 combinations. These can be thought of as odds ratios obtained from cross-tabulations including only two relevant son occupations and two relevant ancestor occupations. In practice, the coefficients are estimated jointly using a multinomial logit model with age controls.

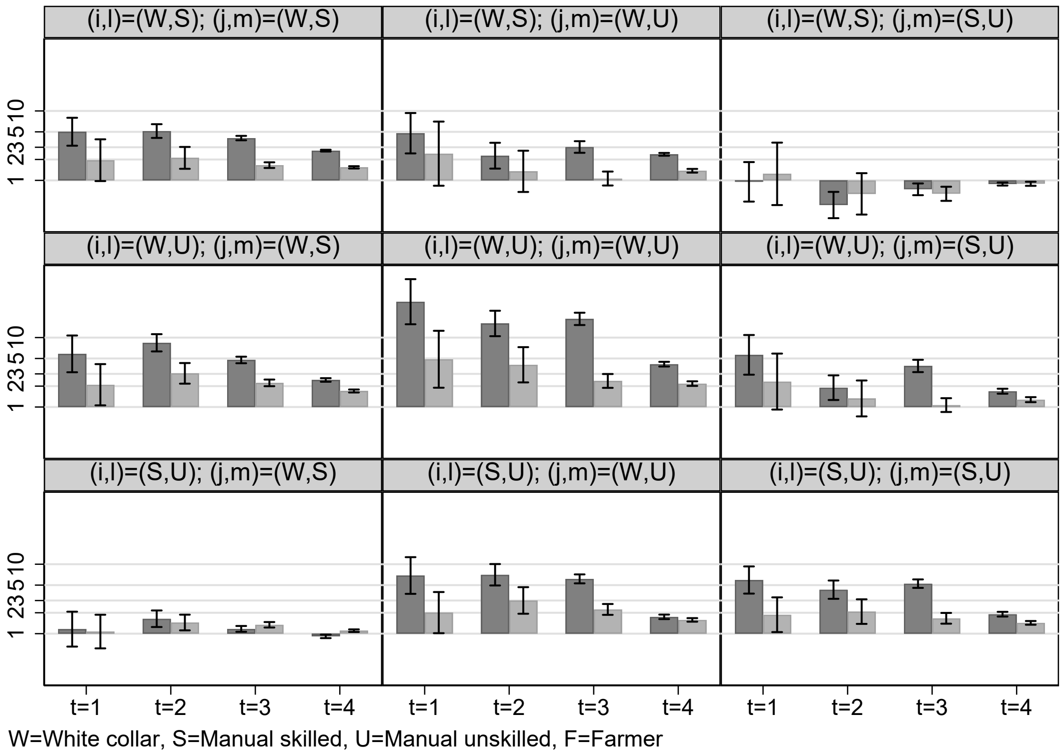

Figure A1: Odds ratios (parameters in logit regression) from $2 \times 2 \times 2$ subtables where (i, l), shown in rows, refers to father's and grandfather's occupation and (j, m), shown in columns, refers to son's occupation. Left (blue) bars denote coefficient on father while right (red) bars denote coefficient on grandfather.

A graphical overview of the coefficient on father's and grandfather's occupation — analogous to odds ratios in 2×2 tables — is given in Figure A1, where the bars denote 95% confidence intervals. Comparisons involving farmers are not shown in the figure, reducing the number of subpanels from 36 to 9. The diagonal shows comparisons of similar occupational pairs for sons and ancestors. In these cases, the parameters are of high magnitude; the largest coefficients are found in the middle panel, where the right bars denote the excess odds of a son entering a white collar occupation rather than an unskilled occupation given that his grandfather held a white-collar occupation rather than an unskilled occupation, for given father's occupation. These are above 2 in all time periods.

The cross-terms comparing white-collar occupations to something else for both sons and ancestors are generally similar to those on the diagonal. For example, having ancestors with white-collar occupations over manual skilled occupations increases the likelihood of entering white-collar occupations over unskilled occupations. However, other terms are very small; the cross terms comparing white collar to manual skilled for sons for manual skilled and manual skilled ancestors are below 1. In sample C, for a father with a given manual occupation, for the final generation it is more likely to enter a white-collar occupation if the grandfather held a manual *unskilled* occupation than if he held a manual *skilled* occupation. This reflects persistence within the manual skilled occupational group. For odds ratios comparing farmers to non-farmers, either on the son or ancestor side, the magnitudes of the odds ratios are generally larger.

Figures A2-A4 show the remaining 27 two-way odds ratios. Associations involving farmers are generally stronger. We observe some cross-occupational effects. For example, as shown in the top right panel of Figure A3; in the first time period, grandsons of skilled workers relative to unskilled workers had higher probability of entering white-collar occupations relative to farm occupations. Nonetheless, the results found here are consistent with those obtained using the simpler set of 2×2 tables in the rest of the paper.

One could further compare the probabilities of outcomes for sons contingent on different pairs of occupations for fathers and grandfathers. For the $4 \times 4 \times 4$ tables used here, there are a total of $(4 \cdot 3/2)^3 = 216$ unique such odds ratios, some of which will be sensitive to very low observation counts. A manual investigation of these does not give any substantial insight beyond what is described above. For this reason, we now move to summary measures incorporating odds ratios for similar fathers' and grandfathers' occupations.

One way of summarizing odds ratios from tables larger than 2×2 is the Altham statistic (Altham, 1970), used by Long & Ferrie (2013) to take into account "off-diagonal" probabilities. This statistic (denoted d below) is effectively a constant multiplied by the geometric average of all possible log odds ratios in the mobility matrix. Extending the approach used in Equation (5) to a multinomial logit model with three equations, we can use the parameters for father's occupation to construct an Altham statistic and a corresponding confidence interval. Let β_j^i denote the coefficient on the dummy variable for father's occupation i in the equation for son's occupation j. We can then express the Altham statistic d for the father-son associations as (see Modalsli, 2015):

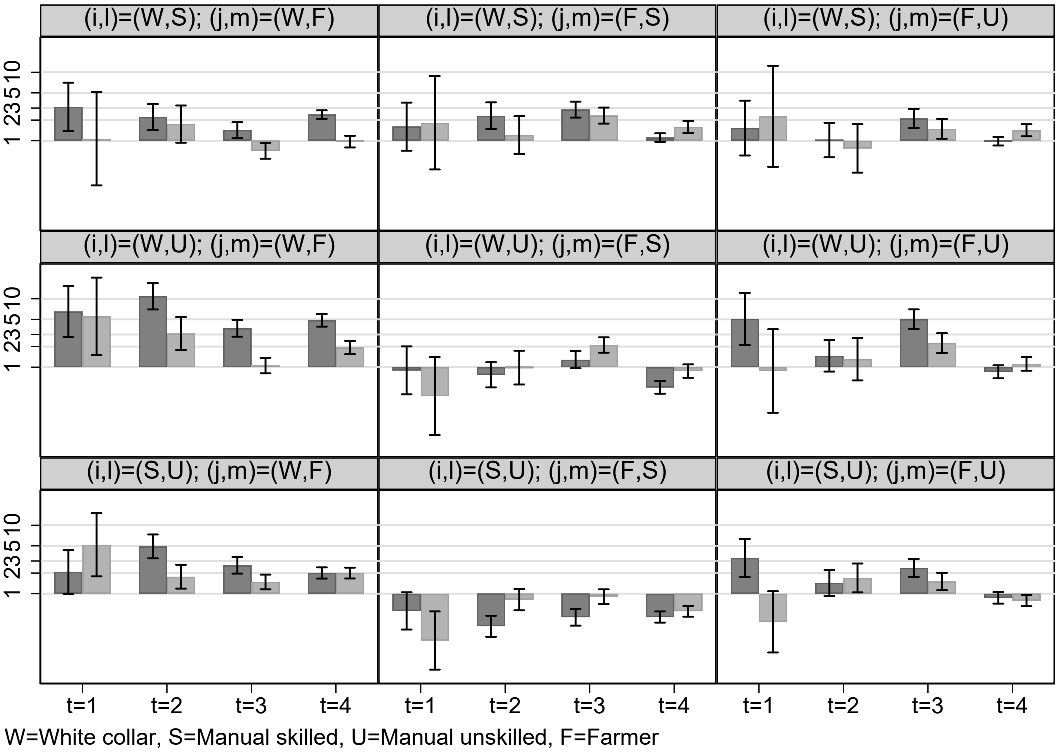

Figure A2: Odds ratios from 2×2 subtables (cf. Figure A1), nonfarm-farm

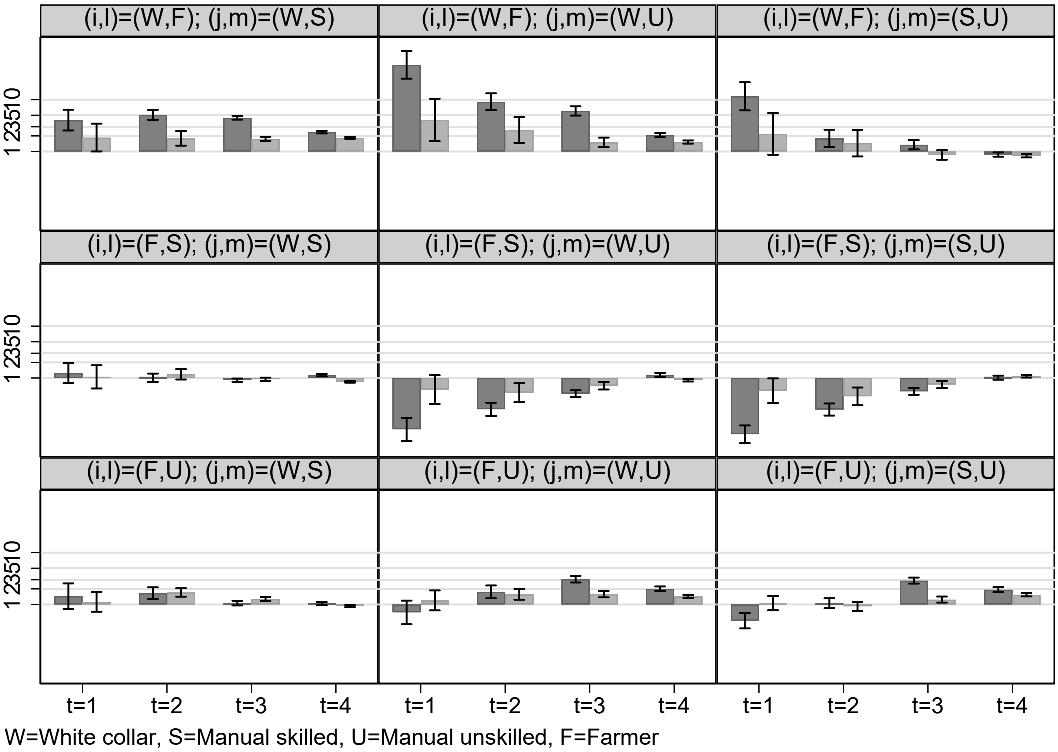

Figure A3: Odds ratios from 2×2 subtables (cf. Figure A1), farm-nonfarm

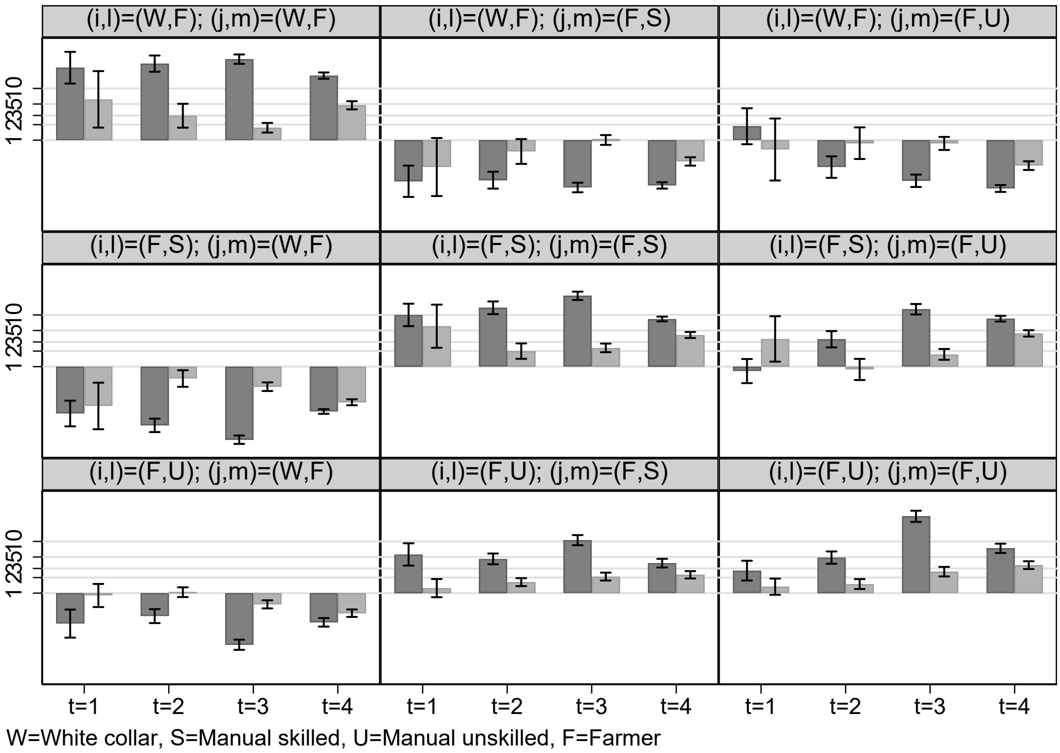

Figure A4: Odds ratios from 2×2 subtables (cf. Figure A1), farm-farm

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

A high value of d corresponds to high odds ratios; that is, low intergenerational mobility.

The statistic can be extended to joint models of father's and grandfather's occupation by including grandfather coefficients in the regression. Table A10 reports Altham statistics for three separate models. First, we model son's occupation by either father or grandfather's occupation separately, with controls only for the ages of each generation in the model. Second, we use a joint model with dummy variables for both father's and grandfather's occupation. In both cases, the statistic reported for grandfathers is obtained by replacing β with γ in Equation (1).

	(1)	(2)	(3)	(4)
Sample	A 1865-1910	B 1865-1960	C 1910-1980	D 1960-2011
Separate models				
Father and son	23.3***	20.1***	22.6***	19.0***
	(22.5 − 24.3)	(19.7 − 20.4)	(22.2 − 23.0)	(18.7 − 19.2)
Grandfather and grandson	21.9***	15.4***	14.6***	14.1***
	(18.9 − 26.2)	(14.0 − 17.2)	(14.1 − 15.1)	(13.6 − 14.6)
Joint model				
Father	20.5***	18.3***	21.0***	14.2***
	(18.4 − 23.3)	(17.1 − 19.7)	(20.3 − 21.7)	(13.6 − 14.8)
Grandfather	11.6***	7.5***	6.4***	8.6***
	(8.2 − 17.7)	(6.3 − 9.8)	(5.7 − 7.4)	(8.1 − 9.3)

Table A10: Father-son and grandfather-grandson Altham statistic calculated using multinomial logit. Higher number reflects higher persistence. Age controls added in all cases. *** indicates 99% significance using χ^2-tests; numbers in parentheses indicate 95% bootstrapped confidence intervals

The first line in Table A10 shows the Altham statistic on father-son mobility, exhibiting a slight decrease (corresponding to increasing mobility) between the first and the final sample. The second line shows the similar statistic for grandfather and grandson. In this case, there is a larger difference between the first and final sample. However, we are primarily interested in the Altham statistics constructed from regressions where both father and grandfather is included at the same time. These are shown in the third and fourth line of Table A10.

We see that the Altham statistic on grandparental occupations is statistically significant even when estimated jointly with father's occupation. Moreover, while the separate models show unambigous increase in mobility between samples C (1910-1960-1980) and D (1960-1980-2011), this can now be interpreted as a substantial increase in father-son intergenerational mobility — a decrease in the influence of fathers — together with a slight *increase* in the influence of grandfathers. However, in the final sample the odds ratios for farmers are high and substantially influence the aggregate Altham statistic even though the farm population is very small.

B.2 Multigenerational income persistence

In the main paper we concern ourselves with the association between father's and grandfather's occupation and the occupation of the son. However, for recent generations, we also have comprehensive income information from the tax registries.

From 1967 onwards the individual records can be linked to tax return registries, where two types of income are recorded. Labor income (*pensjonsgivende inntekt*) is the preferred measure for working-age men, as it reflects the return to occupations. However, for earlier cohorts we must rely on incomes of older men, where a large proportion will have retired. For those 59 years or older we hence use total income (*alminnelig inntekt*), which includes pensions but also capital income. Observations are averaged over five years to remove short-term variations in income, and income is measured at the same age range for all individuals in a given analysis. To abstract from variations in the income distribution over time, we follow Chetty et al (2013) and use the income rank rather than the level of income. Ranks are measured compared to other individuals in the same cohort. Denoting income rank by R and sons, fathers and grandfathers by s, f and g, respectively, the baseline two-generation relationship is

$$R_\iota^s = \alpha + \beta R_\iota^f + \gamma R_\iota^g + \epsilon_\iota \tag{2}$$

For the 1960-1980-2011 sample we can estimate relation (2) directly. For the 1910-1960-1980 sample we do not observe grandfather's income, and instead rely on the grandfather occupation variable to examine the multigenerational process. In that case, the relation becomes

$$R_\iota^s = \alpha + \beta R_\iota^f + \psi \boldsymbol{X}_\iota^g + \epsilon_\iota \tag{3}$$

The relationship (2) is estimated using ordinary least squares, and the results from these rank-rank regressions are given in Table A11. The first column gives the father-son rank correlations for sample C, where sons are born between 1920 and 1950 and incomes are measured at rather advanced ages. The constant term is 38 and the slope term 0.26, meaning that an individual whose father had income at the 25th percentile can be expected to have an income at the $38 + 0.26 \cdot 25 = 44.5$th percentile.

The second column adds controls for grandfather's occupational group, with white-collar as reference group. The slope coefficient is comparable to that in column 1, but there is a substantial difference between those with white-collar grandfathers and other groups. A son whose father had income at the 25th percentile and a white-collar grandfather has an expected income rank of $44+0.24\cdot 25 = 50$, while a son with a father at the same income percentile but a grandfather with a manual skilled occupation has an expected income rank 5.7 percentage points lower. Table A12 shows that the results are robust to measurement of son's income at earlier ages.

Interpreting the results from this mid-twentieth century sample is challenging as we do not

	(1)	(2)	(3)	(4)	(5)	(6)
	Sample C Sons born 1920-1950		Sample D Sons born 1960-1981			
Dependent variable:	Income rank (age 63-67)		Income rank (age 28-32)			
Father income rank (age 63-67)	0.263***	0.243***				
	(42.43)	(36.77)				
Father's income rank (age 28-32)			0.137***	0.135***	0.126***	0.127***
			(46.04)	(44.46)	(40.94)	(41.05)
Grandfather's income rank (age 59-63)					0.0399***	0.0434***
					(13.20)	(13.11)
Grandfather's occ: Farmer		-4.899***		-0.801***		0.521*
		(-8.57)		(-3.14)		(1.90)
Grandfather's occ: Manual skilled		-5.752***		-1.129***		-0.258
		(-9.44)		(-5.22)		(-1.14)
Grandfather's occ: Manual unskilled		-8.748***		-0.458		0.988***
		(-12.29)		(-1.50)		(3.05)
Constant	38.48***	44.49***	48.31***	49.14***	46.83***	46.48***
	(102.83)	(65.12)	(272.53)	(189.22)	(223.52)	(141.18)
N	23700	23700	104555	104555	104555	104555

t statistics in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Table A11: OLS regression on income ranks, samples C (sons born 1920-1950) and D (sons born 1960-1981).

have the same type of information for all generations. However, this is less of a problem in sample D, where the final generation is born between 1960 and 1981. As a baseline specification, we measure both sons' and father's incomes early in their careers. Columns 3 and 4 correspond to the two analyses of sample C above. The rank-rank coefficient is substantially smaller than in sample C, showing that the relationship between father's and son's income is weaker. The coefficients on grandparental occupations are also smaller in magnitude. However, we cannot say whether this truly reflects lower income persistence as incomes are measured at different ages in the two samples.

Column 5 gives the basic rank-rank specification for three generations, with grandparental income measured as an average for ages 59-63. There is a small but significant coefficient on grandparental income rank. An individual whose father and grandfather were both at the 10th percentile would have an expected income rank of $47 + 0.13 \cdot 10 + 0.04 \cdot 10 = 48.7$, while an individual with a father at the 10th percentile but a grandfather at the 90th percentile would have an expected income rank $0.04 \cdot (90 - 10) = 3.2$ percentage points higher. Column 6 indicates that the effect of grandfather's occupation is not only reflected in income, as some of the dummy variables still have significant coefficient values. Table A13 shows that the results are robust to measurement of father's income at different ages.

Bratberg *et al.* (2017) find that rank-rank curves slope upward at the top in Norway and

Sweden. This suggests a separate "top income" effect. A simple way to control for this is to add dummy variables indicating whether father's and grandfather's income is in the top 10 percent. The results of this exercise are reported in Table A14. While the upward slope of the father-son rank correlation is replicated, with an excess rank from top 10 of 8.4 for sample C and 1 for sample D, the coefficient on grandfathers is small and not significant. While we cannot rule out such a top income effect, the evidence here is not strong.

Measurement error in father's or grandfather's income can affect the estimation results for multigenerational persistence in incomes. We can assess the potential for measurement error by examining how the grandfather coefficient changes when the measurement of parental income is improved.[1] We should keep in mind that the potential for bias is already reduced by taking the average of incomes over several years. Adding a squared term for father's income rank does not change the parameter for grandfather's income rank. Adding control for mother's income rank in addition to that for father reduces the grandparental coefficent from 0.040 to 0.036.

These results for multigenerational income persistence are consistent with what was found for occupational categories. As income is one-dimensional and income data is not available for all time periods, it is not possible to look for changes over time similar to those in the occupational data, such as lower persistence in white-collar occupations or changes in manual occupational groups. However, the absence of a "top income" effect for grandfathers suggests that white-collar persistence is not merely an elite phenomenon, but also reflects dynamics further down in the income distribution.

Supplemental tables

Tables A12-A13 show that adjusting the age intervals or covariates does not change the conclusions in Section B.2.

Rank-rank with top income dummies: We estimate the two- and three-generation equations

$$R_{2,\iota} = \alpha + \beta R_{1,\iota} + \phi 1(R_{1,\iota} > 0.9) + \epsilon_\iota \tag{4}$$

$$R_{2,\iota} = \alpha + \beta R_{1,\iota} + \phi 1(R_{1,\iota} > 0.9) + \gamma R_{0,\iota} + \psi 1(R_{0,\iota} > 0.9) + \epsilon_\iota \tag{5}$$

[1] A more comprehensive treatment of bias in multigenerational income regressions in general is provided by Modalsli & Vosters (2019).

	(1)	(2)	(3)	(4)	(5)
Sample:			C (Sons born 1920-1950)		
Dependent variable:		Income rank (R), age 63-67		R, age 59-63	R, age 35-39
Father's income rank (age 63-67)	0.263***	0.243***	0.205***	0.206***	0.220***
	(42.43)	(36.77)	(28.11)	(28.84)	(31.64)
Father's occ: Farmer			-6.996***	-7.008***	-12.14***
			(-11.51)	(-11.77)	(-20.94)
Father's occ: Manual skilled			-6.442***	-5.931***	-6.416***
			(-12.23)	(-11.49)	(-12.78)
Father's occ: Manual unskilled			-6.828***	-7.086***	-10.23***
			(-9.11)	(-9.65)	(-14.33)
Grandfather's occ: Farmer		-4.899***	-2.129***	-2.425***	-5.540***
		(-8.57)	(-3.47)	(-4.03)	(-9.50)
Grandfather's occ: Manual skilled		-5.752***	-3.730***	-3.455***	-3.281***
		(-9.44)	(-5.89)	(-5.57)	(-5.45)
Grandfather's occ: Manual unskilled		-8.748***	-6.122***	-6.317***	-7.785***
		(-12.29)	(-8.24)	(-8.69)	(-11.05)
Constant	38.48***	44.49***	49.42***	49.39***	50.03***
	(102.83)	(65.12)	(63.32)	(64.62)	(67.55)
N	23700	23700	23700	24371	25016

t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A12: OLS regression on income ranks, various income definitions, sample C. Dependent variable: son's income rank

	(1)	(2)	(3)	(4)	(5)	(6)
Sample:			D (Sons born 1960-1981)			
Dependent variable:			Income rank (R), age 28-32			
Father's income rank (age 28-32)	0.135***	0.134***	0.124***	0.126***	0.119***	
	(49.88)	(48.11)	(43.32)	(40.94)	(37.39)	
Father's income rank (age 59-63)						0.135***
						(42.38)
Grandfather's income rank (age 59-63)				0.0399***	0.0401***	0.0425***
				(13.20)	(12.03)	(12.75)
Father's occ: Farmer			-3.555***		-3.653***	-3.422***
			(-10.32)		(-9.47)	(-9.00)
Father's occ: Manual skilled			-2.397***		-2.089***	-0.897***
			(-14.22)		(-11.36)	(-4.73)
Father's occ: Manual unskilled			-1.849***		-1.716***	-1.086***
			(-5.06)		(-4.26)	(-2.68)
Grandfather's occ: Farmer		-0.826***	0.369		1.568***	1.167***
		(-3.55)	(1.48)		(5.40)	(4.01)
Grandfather's occ: Manual skilled		-1.214***	-0.521**		0.269	0.566**
		(-6.09)	(-2.54)		(1.17)	(2.44)
Grandfather's occ: Manual unskilled		-0.386	0.461		1.623***	1.458***
		(-1.39)	(1.60)		(4.89)	(4.38)
Constant	48.35***	49.20***	50.45***	46.83***	47.82***	46.28***
	(301.09)	(207.40)	(201.10)	(223.52)	(137.93)	(129.84)
N	124302	124302	124302	104555	104555	103007

t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A13: OLS regression on income ranks, various income definitions, sample D. Dependent variable: son's income rank

	(1)	(2)	(3)
Sample:	C	D	D
Dependent variable:	R, avg. age 63-67	R, avg. age 28-32	R, avg. age 28-32
Father's income rank (age 63-67)	0.204***		
	(26.91)		
— in top 10	8.411***		
	(13.18)		
Father's income rank (age 28-32)		0.131***	0.121***
		(37.12)	(33.70)
— in top 10		1.004***	0.812***
		(3.24)	(2.61)
Grandfather's income rank (age 59-63)			0.0386***
			(10.93)
— in top 10			0.164
			(0.52)
Constant	40.43***	48.53***	47.06***
	(100.76)	(256.41)	(204.50)
N	23700	104555	104555

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A14: Rank-rank with top income dummies

B.3 Grandparental presence

Measuring presence using a continuous rather than dichotomous variable:

To examine the association between geographic distance and grandparental influence, we introduce a variable Δ denoting the distance between grandfather's residential municipality (at the time his occupation is observed) and grandson's residential municipality in his childhood (at the time father's occupation is observed). This variable is then interacted with the dummy for grandfather's occupation:

$$\log\left(\frac{\Pr(\text{Son's occ} = Z)_\iota}{\Pr(\text{Son's occ} \neq Z)_\iota}\right) = \alpha + \beta D_\iota^f + \gamma_0 \Delta_\iota^g + \gamma_1 D_\iota^g + \gamma_2 \Delta_\iota^g D_\iota^g + \sum_{q \in \{s,f,g\}} \left(\delta \cdot age_\iota^q + \zeta \cdot (age_\iota^q)^2\right) + \epsilon_\iota \tag{6}$$

The distance Δ is measured in units of 100 km. Around one in four dynasties changes location between the two observations; the longest distance moved is 1600 km. The coefficient on the interaction term (γ_2) is presented in the first panel of Table A15.

For white-collar workers, the relationship with distance has the expected sign: the point estimate in the first period of 0.989 indicates a 1.1 percentage point lower association with grandfather's occupation if the grandfather lived 100 kilometres farther away. However, the relationship is weak and not statistically significant. Again, farmers are different with strong and significant coefficients, indicating the special role of this occupational group and the strong relationship between a farmer and a specific farm. The coefficient on manual skilled and manual unskilled is slightly stronger in the first period, but small and insignificant in later periods.

The full set of regression coefficients, as well as regressions on a sample restricted to only those who move, is presented below (Tables A19-A20). On balance, there is no strong evidence that direct interactions drive the association between outcomes across generations beyond father and child; however, some direct effect cannot be ruled out. The results are consistent with those found (for shorter geographical distances) by Knigge (2016) for 19th-century Netherlands.

Moving to the analysis of presence where we rather consider the year of death of the grandfather, we can similarly use a continuous variable measuring the childhood years in which the grandfather and the grandson were both alive. This will vary from zero (when the grandfather dies before the grandson is born) to 16 years (chosen as a reasonable value for the end of the upbringing of the child). This setup is equivalent to that shown in Equation (6), and the coefficient γ_2 is reported in the second panel of Table A15.

The analysis does not support any strong association between multigenerational persistence and "exposure time" measured in this way. The point estimate for white collar workers is 1.002: a 0.2 percentage point higher association with grandfather's occupation for each extra year the grandfather was alive. The 95% interval for this coefficient is $(0.997, 1.007)$, or from 0.3 percentage point lower to 0.7 percentage point higher association.

Detailed tables of regression results for the analysis in the main text:

	(1)	(2)	(3)	(4)
Sample	A	B	C	D
	1865-1910	1865-1960	1910-1980	1960-2011
	Distance moved (in 100 km)			
White collar	0.989	0.908	0.977	0.986
	(-0.05)	(-1.26)	(-1.01)	(-1.64)
Farmer	0.533**	0.933	0.874**	0.790***
	(-2.11)	(-0.86)	(-2.01)	(-3.00)
Manual skilled	0.691	0.857	1.000	1.020**
	(-1.08)	(-1.32)	(0.01)	(2.57)
Manual unskilled	0.792	1.067	1.019	0.985
	(-1.19)	(0.74)	(0.38)	(-1.04)
	Years (in youth) in which grandfather was alive			
White collar				1.002
				(0.75)
Farmer				0.995
				(-0.58)
Manual skilled				1.001
				(0.24)
Manual unskilled				1.002
				(0.41)
Son observed:	1910	1960	1980	2011
Father observed:	1900	1910	1960	1980
Grandfather observed:	1865	1865	1910	1960

Table A15: Distance moved / mortality and grandparental influence. The outcome is grandson's occupation. The coefficient shown (from (6)) is the interaction between grandparental occupation and one of two other characteristics of the dynasty: geographical distance (Panel 1) and years during grandson's childhood when grandfather was alive (Panel 2). Separate logit regressions for each occupational category and sample.

See Section 5 of the main paper. Tables A16-A18 show the model with more detailed specification of the parent generation, separately for movers and mortality, while Table A21 shows the corresponding relationship for income ranks.

	(1)	(2)	(3)	(4)
Sample	A	B	C	D
	1865-1910	1865-1960	1910-1980	1960-2011
	Occupation: White collar			
Non-movers	3.432**	2.397***	1.508***	1.450***
	(2.081)	(4.076)	(6.976)	(16.109)
Movers	1.896	2.529***	1.281***	1.375***
	(0.676)	(3.298)	(3.043)	(8.737)
Difference	1.810	0.948	1.177	1.055
	(0.532)	(-0.151)	(1.624)	(1.235)
	Occupation: Farmer			
Non-movers	1.762***	1.638***	2.133***	3.199***
	(2.648)	(5.472)	(9.728)	(17.076)
Movers	0.901	1.576***	1.156	1.975***
	(-0.193)	(2.674)	(1.012)	(3.844)
Difference	1.956	1.039	1.845***	1.620**
	(1.152)	(0.200)	(3.757)	(2.542)
	Occupation: Manual, skilled			
Non-movers	2.941***	1.283*	1.083*	1.081***
	(2.804)	(1.711)	(1.889)	(4.422)
Movers	0.915	1.214	1.128*	1.101***
	(-0.122)	(1.002)	(1.705)	(2.947)
Difference	3.215	1.057	0.960	0.982
	(1.410)	(0.228)	(-0.500)	(-0.486)
	Occupation: Manual, unskilled			
Non-movers	1.209	1.461***	1.513***	1.349***
	(0.980)	(3.495)	(5.397)	(7.772)
Movers	0.967	1.335	1.959***	1.433***
	(-0.084)	(1.463)	(3.969)	(5.005)
Difference	1.251	1.094	0.772	0.942
	(0.500)	(0.399)	(-1.391)	(-0.735)
Son observed:	1910	1960	1980	2011
Father observed:	1900	1910	1960	1980
Grandfather observed:	1865	1865	1910	1960

Table A16: Odds ratio coefficients on grandfather's occupation, separate regressions for movers (between observation of grandfather and grandson) and non-movers. "Difference" is the linear difference between parameters (log odds ratios), i.e. the ratio of the two displayed coefficients. In contrast to Table 7, the regressions here are run with a full set of dummies for mother and father from Table 6.

Grandfather's occ: 1960	Sample D $\tau = 1980$	Sample D $\tau = 2011$
Occupation: White collar		
Grandfather not alive at τ	1.677***	1.630***
	(14.975)	(29.472)
Grandfather alive at τ	1.618***	1.652***
	(26.290)	(6.708)
Difference	1.037	0.986
	(0.918)	(-0.179)
Occupation: Farmer		
Grandfather not alive at τ	3.319***	3.856***
	(10.038)	(24.477)
Grandfather alive at τ	4.053***	6.518***
	(22.944)	(5.986)
Difference	0.819	0.592
	(-1.488)	(-1.651)
Occupation: Manual, skilled		
Grandfather not alive at τ	1.037	1.017
	(1.333)	(1.346)
Grandfather alive at τ	1.018	1.105
	(1.221)	(1.610)
Difference	1.019	0.921
	(0.614)	(-1.301)
Occupation: Manual, unskilled		
Grandfather not alive at τ	1.713***	1.648***
	(9.749)	(18.237)
Grandfather alive at τ	1.633***	1.758***
	(15.988)	(4.292)
Difference	1.049	0.937
	(0.754)	(-0.485)

Table A17: Mobility and grandparental mortality, sample D; two definitions of grandfather's survival.

Grandfather's occ: 1960	Sample D $\tau = 1980$	$\tau = 2011$
Occupation: White collar		
Grandfather not alive at τ	1.455***	1.437***
	(9.323)	(18.355)
Grandfather alive at τ	1.435***	1.569***
	(16.249)	(4.117)
Difference	1.014	0.916
	(0.305)	(-0.788)
Occupation: Farmer		
Grandfather not alive at τ	2.586***	3.026***
	(6.967)	(17.485)
Grandfather alive at τ	3.175***	6.969***
	(16.355)	(4.421)
Difference	0.815	0.434
	(-1.335)	(-1.881)
Occupation: Manual, skilled		
Grandfather not alive at τ	1.132***	1.092***
	(3.828)	(5.611)
Grandfather alive at τ	1.080***	1.107
	(4.356)	(1.140)
Difference	1.049	0.987
	(1.286)	(-0.147)
Occupation: Manual, unskilled		
Grandfather not alive at τ	1.378***	†
	(4.676)	
Grandfather alive at τ	1.355***	1.690***
	(7.787)	(2.596)
Difference	1.017	†
	(0.209)	

Table A18: Mobility and grandparental mortality, sample D; two definitions of grandfather's survival. In contrast to Table A17, the regressions here are run with a full set of dummies for mother and father from Table 6. † denotes regressions in which the ML procedure did not converge.

Table A19 shows the full results of Regression (6) on grandparental presence, with Δ=distance moved (in 100s of km) in the first four columns and the years (< 16) in which both grandfather and grandson were alive in the fifth column. Table A20 shows the same relationship for the intensive margin, that is, with zeroes excluded.

Presence and income persistence

For the final samples, the variation of multigenerational income persistence with geographical moves and grandfather mortality can be investigated. Again, income ranks yield estimates consistent with those obtained using occupational groups. We recall from Table A11 that the coefficient on grandfather's income in the full 1960-2011 sample is 0.040. Once again splitting the sample according to grandfather's location, the coefficient on grandfather's income is 0.048 for non-movers and 0.030 for movers. In other words, there is substantial grandparental persistence both in cases where individuals move away from their origin and in cases where they stay in the same place, and there is a statistically significant but small difference between the subsamples.

Splitting the sample by mortality yields a coefficient of 0.054 in the subsample where the grandfather survived to the time of the grandson's income measurement (age 32) and 0.038 when he did not. The difference of 0.016 is not statistically significant.[2] Overall, the rank-rank income regressions confirm the results of the regressions using occupational categories; persistence is somewhat amplified by the physical presence of the grandfather, but not greatly so.

[2]The results of the subsample analyses on income ranks are given in Table A21.

Grandfather-grandson treatment:	Geographical distance (in 100 km)				Years (< 16) both alive
	(1)	(2)	(3)	(4)	(5)
Sample	A	B	C	D	D
	1865-1910	1865-1960	1910-1980	1960-2011	1960-2011
Occupation: White collar					
Father same occupation	11.61***	7.635***	5.019***	2.693***	2.729***
	(13.40)	(22.35)	(47.12)	(78.12)	(79.41)
Grandfather same occupation	2.799***	2.643***	1.775***	1.636***	1.618***
	(3.21)	(5.99)	(12.65)	(28.71)	(24.51)
Grandfather treatment level	1.066	1.112***	1.087***	1.051***	0.998
	(0.67)	(3.71)	(7.08)	(12.86)	(-1.35)
Grandfather treatment interaction	0.989	0.908	0.977	0.986	1.002
	(-0.05)	(-1.26)	(-1.01)	(-1.64)	(0.75)
Occupation: Farmer					
Father same occupation	3.491***	7.986***	18.26***	8.115***	8.635***
	(7.82)	(23.67)	(44.02)	(42.10)	(43.93)
Grandfather same occupation	1.742***	1.493***	1.962***	4.023***	3.965***
	(3.37)	(5.07)	(10.72)	(24.76)	(22.22)
Grandfather treatment level	1.104	0.994	0.988	0.894***	0.996
	(0.75)	(-0.10)	(-0.34)	(-3.37)	(-0.59)
Grandfather treatment interaction	0.533**	0.933	0.874**	0.790***	0.995
	(-2.11)	(-0.86)	(-2.01)	(-3.00)	(-0.58)
Occupation: Manual, skilled					
Father same occupation	5.387***	3.346***	2.365***	2.153***	2.171***
	(12.14)	(17.10)	(32.05)	(61.41)	(62.24)
Grandfather same occupation	2.213***	1.389***	0.953	1.007	1.019
	(3.73)	(3.13)	(-1.50)	(0.57)	(1.26)
Grandfather treatment level	1.100	0.957*	0.904***	0.939***	1.000
	(1.44)	(-1.69)	(-8.86)	(-12.24)	(0.00)
Grandfather treatment interaction	0.691	0.857	1.000	1.020**	1.001
	(-1.08)	(-1.32)	(0.01)	(2.57)	(0.24)
Occupation: Manual, unskilled					
Father same occupation	1.683***	2.345***	5.539***	2.193***	2.204***
	(3.75)	(8.39)	(30.51)	(23.07)	(23.28)
Grandfather same occupation	0.997	1.204**	1.548***	1.677***	1.638***
	(-0.03)	(1.97)	(6.79)	(18.40)	(15.25)
Grandfather treatment level	0.884*	0.898*	0.938**	0.970***	1.006***
	(-1.67)	(-1.96)	(-2.31)	(-4.51)	(3.10)
Grandfather treatment interaction	0.792	1.067	1.019	0.985	1.002
	(-1.19)	(0.74)	(0.38)	(-1.04)	(0.41)
N	2086	6039	28084	131076	131076
Son observed:	1910	1960	1980	2011	2011
Father observed:	1900	1910	1960	1980	1980
Grandfather observed:	1865	1865	1910	1960	1960

Exponentiated coefficients; t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A19: Coefficients from regression (6). “Level” is distance from grandfather (in 100 km) in columns (1)-(4); years grandfather is alive simultaneously with grandson in column (5). “Interaction” is interaction between level and grandfather’s occupation. Extensive margin (includes those with zero for the level variable)

Grandfather-grandson treatment:	Geographical distance (in 100 km)				Years (< 16) both alive
	(1)	(2)	(3)	(4)	(5)
Sample	A	B	C	D	D
	1865-1910	1865-1960	1910-1980	1960-2011	1960-2011
Occupation: White collar					
Father same occupation	9.046***	7.669***	5.179***	2.811***	2.658***
	(6.51)	(14.18)	(26.70)	(41.45)	(50.61)
Grandfather same occupation	1.766	3.526***	1.370***	1.551***	1.665***
	(0.95)	(4.20)	(3.48)	(12.19)	(10.64)
Grandfather treatment level	0.924	1.050	1.013	1.020***	0.999
	(-0.62)	(1.49)	(0.97)	(4.42)	(-0.40)
Grandfather treatment interaction	1.160	0.867	1.034	0.991	1.000
	(0.59)	(-1.64)	(1.24)	(-0.97)	(0.10)
Occupation: Farmer					
Father same occupation	4.800***	13.46***	29.90***	17.51***	9.533***
	(4.01)	(15.12)	(22.89)	(17.12)	(28.51)
Grandfather same occupation	1.207	1.539**	1.341*	2.324***	3.518***
	(0.42)	(2.25)	(1.84)	(4.58)	(8.42)
Grandfather treatment level	1.256	1.078	0.951	0.927**	0.993
	(1.49)	(1.21)	(-1.06)	(-2.03)	(-0.61)
Grandfather treatment interaction	0.747	0.918	0.937	0.949	0.999
	(-1.09)	(-0.95)	(-0.77)	(-0.69)	(-0.04)
Occupation: Manual, skilled					
Father same occupation	5.173***	3.136***	2.912***	2.404***	2.130***
	(6.47)	(10.18)	(19.77)	(33.58)	(39.74)
Grandfather same occupation	2.302	1.326	1.174**	1.121***	1.070*
	(1.56)	(1.27)	(2.13)	(3.57)	(1.78)
Grandfather treatment level	1.072	0.931**	0.974**	0.984***	1.000
	(0.87)	(-2.34)	(-2.05)	(-2.73)	(-0.12)
Grandfather treatment interaction	0.716	0.844	0.960	0.998	0.998
	(-0.76)	(-1.13)	(-1.37)	(-0.26)	(-0.70)
Occupation: Manual, unskilled					
Father same occupation	4.536***	2.533***	6.000***	1.971***	2.107***
	(5.11)	(4.67)	(12.62)	(7.96)	(14.60)
Grandfather same occupation	0.811	1.313	2.143***	1.711***	1.840***
	(-0.63)	(1.26)	(4.29)	(7.64)	(7.67)
Grandfather treatment level	0.909	1.001	1.019	0.999	1.008**
	(-1.04)	(0.02)	(0.62)	(-0.18)	(2.40)
Grandfather treatment interaction	0.881	1.040	0.945	0.980	0.993
	(-0.52)	(0.43)	(-1.02)	(-1.18)	(-0.92)
N	393	1588	6762	34502	56484
Son observed:	1910	1960	1980	2011	2011
Father observed:	1900	1910	1960	1980	1980
Grandfather observed:	1865	1865	1910	1960	1960

Exponentiated coefficients; t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A20: Coefficients from regression (6). "Level" is distance from grandfather (in 100 km) in columns (1)-(4); years grandfather is alive simultaneously with grandson in column (5). "Interaction" is interaction between level and grandfather's occupation. Intensive margin (excludes zeroes)

Sample	Sample D
Non-movers	0.0476***
	(11.34)
Movers	0.0303***
	(7.01)
Difference	0.0172***
	(2.86)
Grandfather alive at son's age 32	0.0538***
	(6.38)
Grandfather not alive at son's age 32	0.0380***
	(11.76)
Difference	0.0158
	(1.74)

Table A21: Rank-rank table with movers and mortality. Dependent variable: son's income rank age 28-32. The coefficient shown is that for grandfather's income rank (age 59-63), controlling for father's income rank (age 28-32).

References

Altham, Patricia M E. 1970. The Measurement of Association of Rows and Columns for an $r \times s$ Contingency Table. *Journal of the Royal Statistical Society Series B*, **32**(1), 63–73.

Blau, Peter M, & Duncan, Otis. 1967. *The American Occupational Structure.* Wiley.

Bratberg, Espen, Davis, Jonathan, Mazumder, Bhashkar, Nybom, Martin, Schnitzlein, Daniel, & Vaage, Kjell. 2017. A Comparison of Intergenerational Mobility Curves in Germany, Norway, Sweden and the U.S. *Scandinavian Journal of Economics*, **119**(1), 72–101.

Knigge, Antonie. 2016. Beyond the Parental Generation: The Influence of Grandfathers and Great-grandfathers on Status Attainment. *Demography*, **53**, 1219–1244.

Long, Jason, & Ferrie, Joseph. 2013. Intergenerational Occupational Mobility in Great Britain and the United States since 1850. *American Economic Review*, **103**(4), 1109–1137.

Mayer, K U. 2007. Retrospective Longitudinal Research: The German Life History Study. *Pages 85–106 of:* Menard, S (ed), *Handbook of Longitudinal Research.* Elsevier.

Modalsli, Jørgen. 2015. Estimating occupational mobility with covariates. *Economics Letters*, **133**, 77–80.

Modalsli, Jørgen. 2017. Intergenerational mobility in Norway, 1865-2011. *Scandinavian Journal of Economics*, **119**(1), 34–71.

Modalsli, Jørgen, & Vosters, Kelly. 2019. Spillover bias in multigenerational income regressions. *Statistics Norway discussion paper*, **897**.