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 VIII 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 IX 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.

VIII Data, matching and robustness analyses

VIII.A Data overview

The data are explained in Sections III.A-III.B 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)

VIII.B 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

VIII.C Great-grandfathers and sample selection

See Section IV.D 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	C	D	D	D	D
	1865-1980	1865-1980	1910-2011	1910-2011	1865-2011	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.

VIII.D 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.

VIII.E 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.

VIII.F 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.

IX Supplemental analyses

IX.A 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.

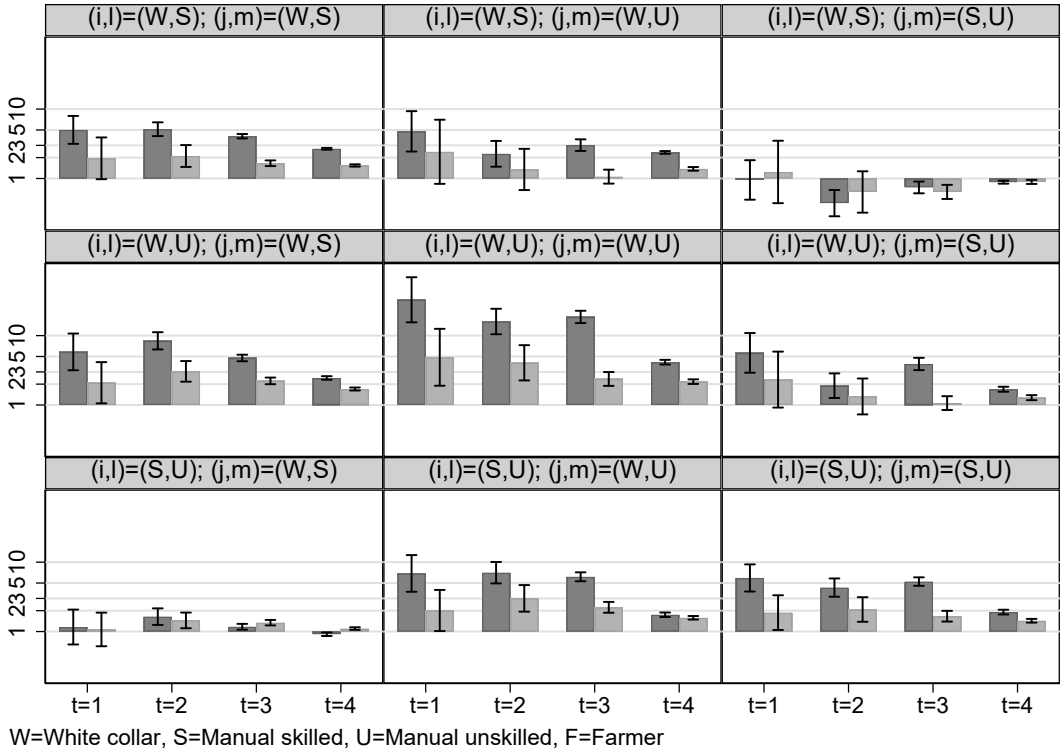

Figure A1: Odds ratios (parameters in logit regression) from $2 \times 2 \times 2$ subtables where (i, l), shown in rows, refers to father's and grandfather's occupation and (j, m), shown in columns, refers to son's occupation. Left (blue) bars denote coefficient on father while right (red) bars denote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

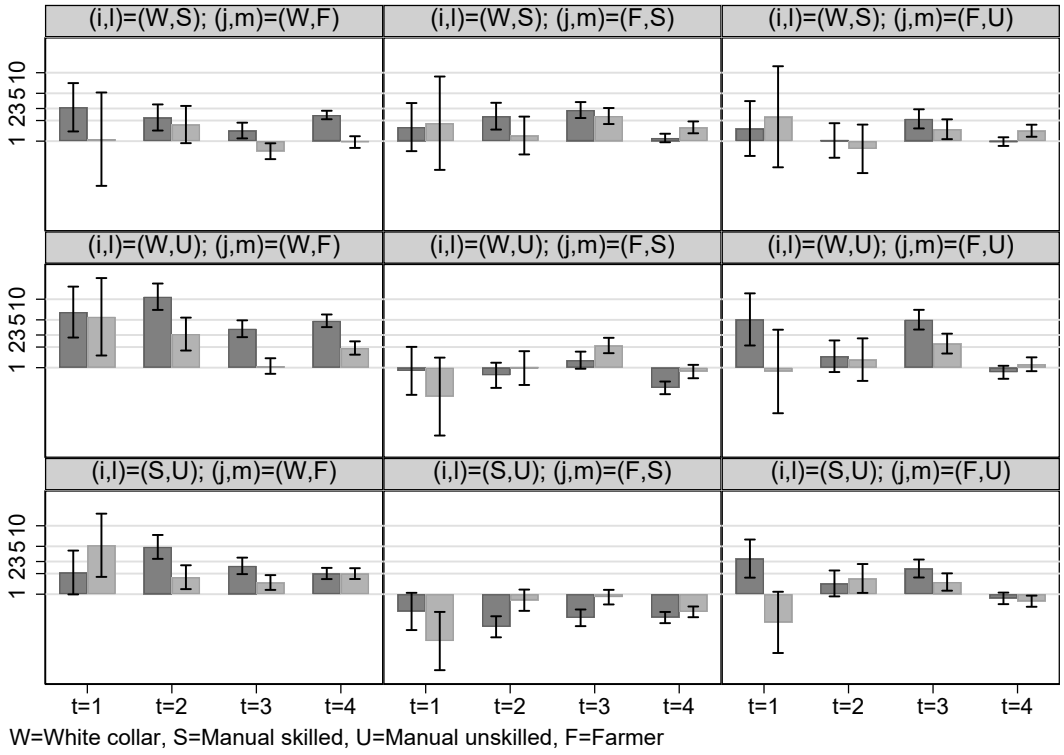

Figure A2: Odds ratios from 2×2 subtables (cf. Figure A1), nonfarm-farm

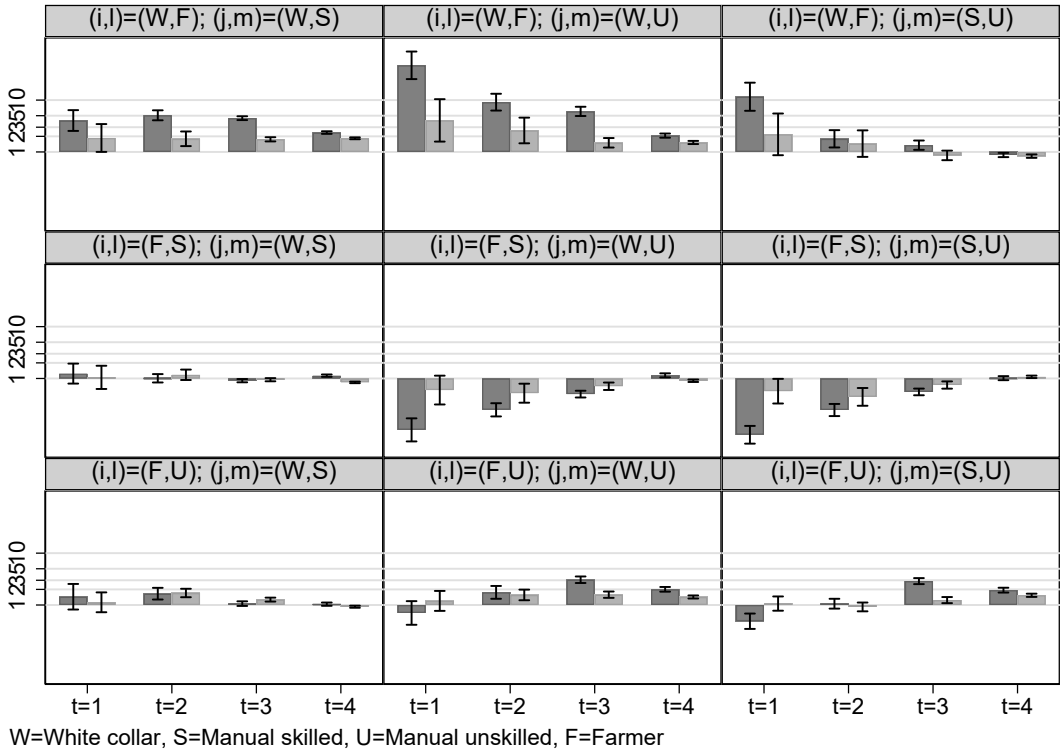

Figure A3: Odds ratios from 2×2 subtables (cf. Figure A1), farm-nonfarm

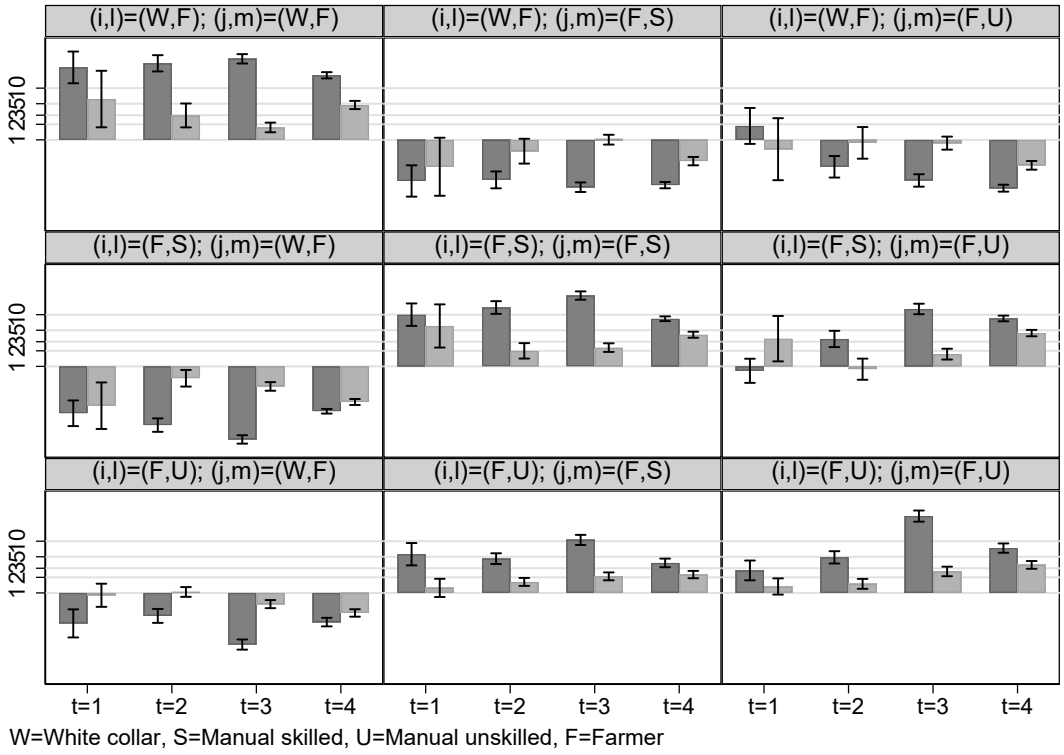

Figure A4: Odds ratios from 2×2 subtables (cf. Figure A1), farm-farm

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

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	B	C	D
	1865-1910	1865-1960	1910-1980	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.

IX.B 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		Sample D			
	Sons born 1920-1950		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 IX.B.

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

IX.C 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 VI 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 $\tau = 1980$	$\tau = 2011$
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**.