Measuring Mobility: Comparing intergenerational status persistence across time and place

Gregory Clark[1*]

Neil Cummins[2]

Matthew Curtis[3]

[1]Department of Economics, University of California, Davis, CA 95616 and Department of Economic History, London School of Economics, UK

[2]Department of Economic History, London School of Economics, UK

[3]ECARES - Université libre de Bruxelles, Brussels, Belgium

Abstract

There has been considerable recent interest in comparing intergenerational social status mobility across time and place. But such attempts are vitiated by unknown measurement errors in status indices, errors that also vary over time and place. Typically the more error, the more seeming social mobility. Using a new database of 1.7 million marriages in England 1837-2021, we show how improving the quality of an occupational status index itself generates lower measured rates of social mobility. Without controlling for the errors embodied in all social status indices, comparing social mobility across time and place is impossible. This paper develops a solution to this problem using not absolute correlations, but the relative correlations between different family members. The method deployed here suggests that occupational status persistence in England 1837-2021 was always much greater than conventionally measured, and was largely unchanging over the long interval 1837-2021.

Introduction

Sociologist and economists have conducted many studies of intergenerational social mobility rates, and have compared these rates across time, and across societies (1-5). A recent study, for example, of intergenerational occupational status correlations for the USA for men born 1830-1980 found the results shown in figure 1 (1). This suggests rapid social mobility in the US throughout the birth years 1830-1980, with intergenerational correlations always less than 0.33. Similarly Long and Ferrie directly estimate occupational mobility for England 1851-1911, using census data, and imputing average earnings for each occupation as a measure of status. They find for 1851-1911 that the intergenerational elasticity of status in England was then 0.27, compared to 0.23 in the USA (2, table 5). But also we see a clear decline in social mobility between mid nineteenth century and the modern period. However, a recent study by Zachary Ward (3) reports a different trend of intergenerational correlations 1830-1980, and much higher intergenerational correlations once measurement error is controlled for.

For England social mobility rates can be measured for the modern period using the CAMSIS occupational status scales, and for 1800-1938 using the HISCAM scales.[1] For this paper we have generated a large new database, which records for 1.7 million church marriages in England 1837-2021 the occupation of the groom, and that of his father and his father-in-law. Using this new data figure 1 shows by decade of birth, 1810 to 1980, for men in England the correlation in occupational status of fathers and sons, using the CAMSIS index from 1991 for the birth decades 1880-1980, and the HISCAM-GB index for the birth decades 1810-1920.

The CAMSIS index used is derived for 371 occupational categories using occupations of husbands and wives in 92,021 marriages in the 1991 census. The HISCAM-GB index is derived for around 400 occupations using 51,419 occupational parings between father and son. These English series suggest much slower rates of social mobility for England across most of the birth decades 1830-1980 than in the USA, but also a trend towards increasing social mobility in England.

---

[1] Both these scales assign occupational status to maximize either the husband-wife status correlation (CAMSIS) or the father-son correlation (HISCAM). For details see https://www.camsis.stir.ac.uk/Data/Britain91.html, https://www.camsis.stir.ac.uk/hiscam/. See also (4-6).

However, with the large set of data we have on nearly 3 million father-son and father-in-law son occupational pairs we are also able to construct a new index for 1800-1939, on the same association principles as HISCAM, but with a different set of 442 occupational categories. We call this the CCC index. As figure 1 shows the CCC index produces even higher intergenerational status correlations for the English birth cohorts 1810-1920. By implication there is even more of an increase in social mobility rates 1810-1980 than with the HISCAM index.

Figure 1 illustrates that a problem with all these measures, and with comparisons of social mobility across time and place, is that the occupational status indices used embody substantial errors. These errors are of two types. First the ascribed status categories will have errors in their status ranking. That error will be smaller the more data is employed to estimate the index, as we see in figure 1. If we had more data for marriages 1940-2021 we could produce a new version of the CAMSIS index which would also show higher father-son status correlations. But second, within each ascribed status category, status can vary by differing amounts with different employment structures in different epochs and countries. This error, however, will not be reduced through more data.

As occupations change over time, the importance of these errors can also change, giving potentially spurious impressions of changes in social mobility rates. Thus for the English marriages 1837-79, the most common occupational descriptor, “labourer,” constituted 27% of grooms, and 30% of their fathers. “Labourer” is a homogeneous work description, with little difference within the category in social status for different types of laborers. For marriages 1980-2021 in contrast the most common occupation for grooms, 10% of occupations, was some kind of manager or supervisor. Manager is a highly heterogenous category, encompassing all kinds of true social status. Thus differences in the structure of occupations over time or across countries can lead to occupational status indices which correspond more or less closely to true social status.

In this paper we show how to measure intergenerational occupational status mobility rates independent of these two types of measurement error. This method employs the ratio of the correlation of status between fathers and fathers-in-law relative to that between son and father-in-law as the measure. Whatever the errors in status attribution to occupations, this ratio will measure the underlying intergenerational correlation in occupational status. If there is rapid intergenerational mobility the correlation of the fathers’ status will be much smaller than the correlation of son to father-in-law. Using our large new marriage database this new measure shows

occupational status mobility rates are substantially lower than conventionally measured. These measures also suggest that in England intergenerational occupational status correlations were close to constant across the birth interval 1810-1989.

Results

Researchers on social mobility have long recognized the potential measurement issues we raised above (7-8). One proposed solution was using abstract measures of movement between social status categories, where social status by category is not parameterized, such as the Altham Index (9-12). But the Altham Index is difficult to interpret as a measure of social mobility (13). Further, any exercise in measuring social mobility that has to assign different occupations or educations to a limited set of social categories, as with the Altham Index, still faces exactly the same problem of differing measurement errors across time and space in assigning these categorizations. Another proposed solution is using additional observations of an individual over time as an instrumental variable (14-15, 3). But this IV strategy, while it deals with the mislabeling of occupational status, will not deal with the measurement problems we outlined above.

Here we show how we can use the marriage records in England to estimate intergenerational occupational correlations that will be independent of both measurement errors discussed above. Since 1837 to the present the marriage certificate in England has recorded the "rank or profession" of grooms, brides and both fathers. Since all these marriage certificates have survived in the archives of the Registrar General there is thus a trove of 110 million marriage records in England 1837-2022 with which potentially to measure intergenerational social mobility. As noted above we have access to a subset of 1.7 million copies of these records from church registers deposited in County Record Offices.

There is also good evidence for England from 1837 onward that matching in marriage was between grooms and brides. Since the average age of first marriage was high throughout this period, being 24 for women and 26 for men, most parties to marriage were legally and financially independent of their parents, with many women also living separately from their birth families as domestic servants. Arranged marriages were uncommon. Marriage to close relatives was also relatively uncommon, with cousin marriages well below 1% of all marriages (16).

While brides often did not have any listed occupation in the earlier English marriage records, we can think of them having a latent occupational status, with which the groom was matching. Their latent status would be their occupational status had they been engaged in the labor market like a man. This latent status of brides was as important as the manifest status of grooms in predicting child outcomes.

Consider figure 2, which shows the pattern of correlations in occupational status between a groom, his father and his father-in-law, assuming that the matching in marriage is between groom and bride. The true correlations in occupational status between father and son, bride and groom, and bride and her father, are assumed to be *b*,*r*, and *f*. The correlation in occupational status between groom and father-in-law and father and father in law will be, with matching between bride and groom, *rbf* and *rf*.

But these correlations, where observed, will be attenuated by measurement errors, measurement errors that vary with time and place. The attenuation will potentially be different where the pair observed is male, θ, as opposed to male and female, ϕ. Figure 2 shows the observed correlations between father and son, father and father-in-law and son and father-in-law. The observed correlation in occupational status between groom and father, and groom and father-in-law will be θb and θrf. The observed correlation in occupational status between father and father-in-law will be θrbf. But this in turn implies that

$$b = \frac{correlation\ father-in-law\ to\ father}{correlation\ father-in-law\ to\ groom} = \frac{\theta rbf}{\theta rf}$$

Thus by taking the ratio of the father-in-law to father and father-in-law to groom correlations we can get an estimate of the underlying intergenerational father-son correlation independent of measurement errors, even when these errors are changing over time periods as in England 1837-2021.

Table 1 shows these intergenerational correlation estimates for father-in-law to groom and fatherin-law to father using the CCC and HISCAM indices for marriages 1837-1979, and the CAMSIS index for marriages 1900-2021.

Despite the measured father-in-law to groom correlation dropping substantially on all the indices between 1837-79 and 1980-2021, the implied intergenerational father-son correlation shown in the table is close to 0.9 all they way from 1837 to 1979, independent of what index is used. Thereafter the point estimate drops to 0.77, though with a standard error now because of fewer observations of 0.03, so that the actual value could be in the range 0.72 to 0.84 in the 5% confidence interval. Importantly the substantial decline in the father to son correlations shown by the three different status indices in figure1 where the decline 1837-2021 is not echoed in the associated estimates of the underlying father-son correlation which goes just from 0.90 down to 0.77.

Note that the underlying marital correlation in underlying social status will be given by

$$r = \frac{correlation\ father-in-law\ to\ groom}{correlation\ to\ groom} * \frac{b}{f} = \frac{\theta rf}{\theta b} * \frac{b}{f}$$

If daughters inherit underlying status as strongly as sons, so that *b* = *f*, then that underlying marital status correlation r will be just

$$r = \frac{correlation\ father-in-law\ to\ groom}{correlation\ to\ groom}$$

There is other evidence suggesting that *b* = *f*, so that there is gender symmetry in the inheritance of (underlying) occupational status (16).

Despite again the declining measured father-son correlations, these implied marital correlations are high, and in this case stable throughout the years 1837-2021, lying always in the range 0.80 to 0.83 whatever status index is used. This very high implied assortment in marriage throughout is consistent with the high intergenerational correlation estimated across these same years (16).

Thus, despite the observed substantial decline over time in father-son, father to father-in-law, and father-in-law to son correlations, the correlation patterns observed are largely consistent with both very high levels of assortment by status in marriage, and a subsequent very strong intergenerational correlation in status. The observed correlations also suggest no change over time in the strength of marital assortment, despite the rise in female education and employment across these years. Intergenerational mobility in status is very limited throughout, with just a modest increase observed in the last period.

Discussion

The average underlying intergenerational correlation in occupational status of 0.88 calculated above is much higher than most existing estimates. There is reason that it may be potentially biased upwards by about 0.05. This correlation is calculated as the ratio of the correlation of status of the fathers to the son-father-in-law correlation. The father-father correlation is calculated a generation earlier than the father-son correlation. If these correlations are declining over time because of changes in the structure of employments, and in the terms used to describe occupations, then potentially they will be biased upwards by this downward drift in measured correlations. Based on the observed data, however, the maximum upward bias for England would be around 0.05. This would imply the true underlying intergenerational correlation would still be 0.83 or greater.

Is such a high intergenerational correlation, so different for example from the recent estimates for the USA 1830-1980 (1) shown in figure 1, plausible? We can show, however, for marriages in England in the period 1837-1879 that the true correlation must be indeed in this range.

The CCC index above shows a measured intergeneration occupational status correlation of 0.70 for marriages 1837-1859. But this measured correlation will still be below the true correlation as a result of three forms of measurement error. The first is the mislabeling of occupations. The second is the mismeasurement of the exact average status of each of the 442 occupation categories. The third is that people whose occupation is assigned to the same of the 442 categories will often actually differ in occupational status. The category "clerk," for example, covers occupations that differ widely in earnings, and in other measures of occupational status.

Just focusing on the last two sources of error, suppose a persons true occupational status is z. Suppose also their assigned status on an occupational index is Z. Then there will be two independent errors linking their assigned status to their true status. $Z = z + u + e$, where e is the error in measuring the true average occupational status of the assigned occupation Z. u is the error caused by the range of occupations that fall under the label Z, each with a different underlying status.

When we measure intergenerational mobility with such a social status index the estimate is biased downwards by a factor

$$\frac{\sigma_z^2}{\sigma_z^2 + \sigma_u^2 + \sigma_e^2}$$

Using an entirely different set of data which links occupations in the period 1800-1939 in England with measures of education and wealth at death we are able to construct another occupational index for this period for England, which we dub here the CCC2 index (17). This index performs almost as well as the CCC index in terms of the

observed intergenerational correlation of status on the marriage database. Yet it has a correlation of just 0.86 in terms of the occupational status assigned to occupation categories.

For the CCC and CCC2 indices, because of their entirely independent construction, the error component *e* attached to errors in the average occupational status by category will be independent, but not the within-category component u. Assuming the error term *e* variance is the same for each of these indices, the correlation between these indices 0.86 will be

$$\rho = \frac{\sigma_z^2 + \sigma_u^2}{\sigma_z^2 + \sigma_u^2 + \sigma_e^2} = 0.86$$

This implies that the error component in these indices we have derived has to be at least 14% of the variance in measured status. It also implies that if we multiply our father-son correlations by 1.16 we will get an estimate closer, but still not as large as, the true underlying persistence of occupational status across generations. Since that correlation for the CCC index for 1837-1859 marriages is 0.70, the true intergenerational correlation in occupational status in this period has to be at least 0.81. When we add the attenuation caused by the mislabeling of occupations, and the variance within occupational categories, the true underlying correlation of occupational status in England 1837-1859 could easily equal the 0.83-0.88 estimated by the new method above. Thus for at least the 1837-1859 period we can verify that an estimated true intergenerational status correlation of 0.83-0.88 is entirely reasonable.

In summary, in this paper we first illustrate the dangers of trying to compare rates of intergenerational mobility across time and place. Measured mobility rates will be strongly influenced by both the quality of the occupational status index, and the characteristics of employment by time period and location.

Next we show how to derive measures of the true underlying father-son occupational status correlation, independent of measurement errors, using the occupational status of fathers, sons and fathers-in-law. These new measures of intergenerational correlations for England are much higher than those measured by conventional estimates. They are in the region of 0.9 for all periods but the most recent, marriages 1980-2021, where it is around 0.78. They suggest that the social world shows far less intergenerational social mobility than has been believed.

There will be resistance to the idea that true social mobility rates could be so low, given the gap between existing estimates of intergenerational correlations, shown in figure 1, and the ones derived here. But we see from the CCC index that for the period 1837-1859 the implied correlation in underlying occupational status father-son has to be substantially greater than 0.81. So at least in this first period we can validate these new estimates.

Materials and Methods

Most of the 1.7 million marriage records used for the estimates here were transcribed by volunteers to the FreeREG organization, and posted on their web page (https://www.freereg.org.uk/), but we collected 32,000 additional marriage records from Essex church records 1837-2021. In particular, for the years 1980-2021 most of the marriage records were from our own supplementary collection.

These marriage records, where the information comes from church registers deposited in local record offices, exclude civil marriages. But though civil marriage was introduced in England in 1837, such marriages remained a small minority of all weddings before 1914. In 1841 Civil marriages were 1.7% of all marriages. In 1914 they were still only 24%, and 31% in 1952 (18).

Because transcribing these marriage records is a volunteer effort based on local interests, the numbers of marriages recorded by county varies considerably. Four counties contain about 50% of the FreeREG marriages transcribed for England: Kent, Lancashire, Lincolnshire, and Staffordshire. But these counties were very different in terms of occupations and urbanization, so that the overall sample generated is representative of England as a whole.

The marriage records for the years 1980-2021 came largely from deposited church marriage registers in Essex Record Office that we collected ourselves. In these years only a minority of all weddings were performed in churches. 49% of weddings by 1982 and 68% by 2012 were civil (18). But there is no reason to expect that the father-son correlations for church weddings would be any different than for the population as a whole.

The new CCC occupational status index for England 1837-1939 employed here was estimated using Goodman's RCII association model (19). This is fitted by iterating to a set of occupational rankings that maximize the correlation of occupational status in father-son and father-in-law-son pairings.

Acknowledgments

We thank Michael Dearing for the funding that allowed the collection of the details of 32,000 marriages from Essex Record Office. Zhiming Zhu of LSE provided excellent RA work in collecting the modern Essex marriage records.

References

1 Song, Xi, Catherine G. Massey, Karen A. Rolf, Joseph P. Ferrie, Jonathan L. Rothbaum, and Yu Xie, "Long-term decline in intergenerational mobility in the United States since the 1850s," *Proceedings of the National Academy of Sciences*, 2020, *117* (1), 251–258.

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

3 Ward, Zachary, "Intergenerational Mobility in American History: Accounting for Race and Measurement Error," NBER Working Paper 29256 September 2021.

4 Lambert, Paul S., Richard L. Zijdeman, Marco H. D. Van Leeuwen, Ineke Maas, and Kenneth Prandy, "The Construction of HISCAM: A Stratification Scale Based on Social Interactions for Historical Comparative Research," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 2013, *46* (2), 77–89.

5 Prandy, Kenneth and Wendy Bottero, "Social Reproduction and Mobility in Britain and Ireland in the Nineteenth and Early Twentieth Centuries," *Sociology*, 2000, 34 (2), 265–281.

6 Prandy, Kenneth and Paul Lambert, "Marriage, Social Distance and the Social Space: An Alternative Derivation and Validation of the Cambridge Scale," *Sociology*, 2003, *37* (3), 397–411.

7 Solon, Gary, "Intergenerational Income Mobility in the United States," *The American Economic Review*, 1992, pp. 393–408.

8 Braun, Sebastian Till and Jan Stuhler, "The Transmission of Inequality Across Multiple Generations: Testing Recent Theories With Evidence From Germany," *The Economic Journal*, March 2018, *128* (609), 576–611.

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

10 Altham, P. M. E. and J. P. Ferrie, "Comparing Contingency Tables Tools for Analyzing Data from Two Groups Cross-Classified by Two Characteristics," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 2007, *40* (1), 3–16.

11 Long, Jason and Joseph Ferrie, "Intergenerational Occupational Mobility in Great Britain and the United States since 1850," *American Economic Review*, June 2013, *103* (4), 1109–37.

12 Perez, Santiago, “Intergenerational Occupational Mobility across Three Continents,” *The Journal of Economic History*, 2019, *79* (2), 383–416.

13 Bouchet-Valat, Milan, “General Marginal-free Association Indices for Contingency Tables: From the Altham Index to the Intrinsic Association Coefficient,” *Sociological Methods & Research*, 2022, *51* (1), 203–236.

14 Altonji, Joseph G. and Thomas A. Dunn, “Relationships Among the Family Incomes and Labor Market Outcomes of Relatives,” June 1991.

15 Modalsli, Jørgen and Kelly Vosters, “Spillover Bias in Multigenerational Income Regressions,” *Journal of Human Resources*, 2019, pp. 0520–10919R2.

16 Clark, Gregory and Neil Cummins, “Assortative Mating and the Industrial Revolution: England, 1754–2021,” CEPR Discussion Paper DP17074, 2022.

17 Clark, Gregory, Neil Cummins, and Matthew Curtis, “The Mismeasure of Man: Why Intergenerational Occupational Mobility Is Much Lower than Conventionally Measured, England, 1800-2021,” Technical Report DP17346, Centre for Economic Policy Research, London, 2022.

18 Haskey, John. “Marriage rites: trends in marriages by manner of solemnisation and denomination in England and Wales, 1841-2012.” In Rebecca Probert, Joanna Miles and Perveez Mody (eds.), *Marriage rites and rights*. London, Hart Publishing, 2015.

19 Goodman, Leo A., “Multiplicative Models for the Analysis of Occupational Mobility Tables and Other Kinds of Cross-Classification Tables,” *American Journal of Sociology*, 1979, 84 (4), 804–819.

Figures and Tables

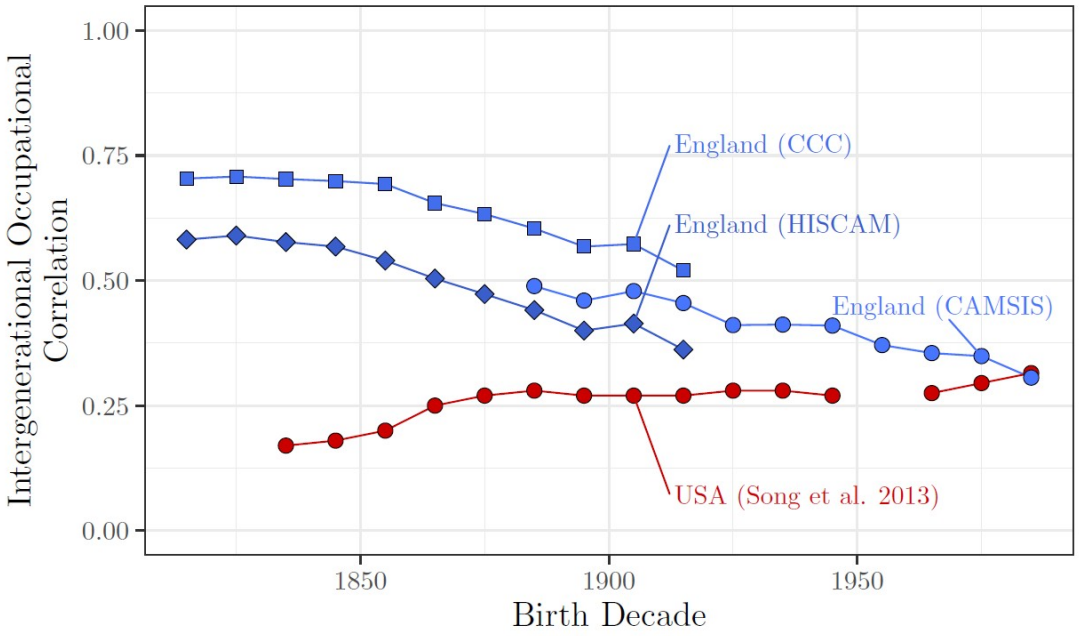

Figure 1. Intergenerational Occupational Status Correlations USA Compared with England, 1810s-1980s

Notes: The figure contrasts estimates of intergenerational occupational persistence for the US from Song et al. (2020) with a set of alternative estimates for England. England (CCC) are the intergenerational occupational correlations 1810-1919, calculated by the new occupational status index presented in this paper using the 1.7m marriage records, England (HISCAM) are the same set of correlation, 1810-1919 but instead using the HISCAM index (Lambert et al. (2013)). England (CAMSIS) uses the CAMSIS index to calculate the correlations 1920-1980 (https://www.camsis.stir.ac.uk/).

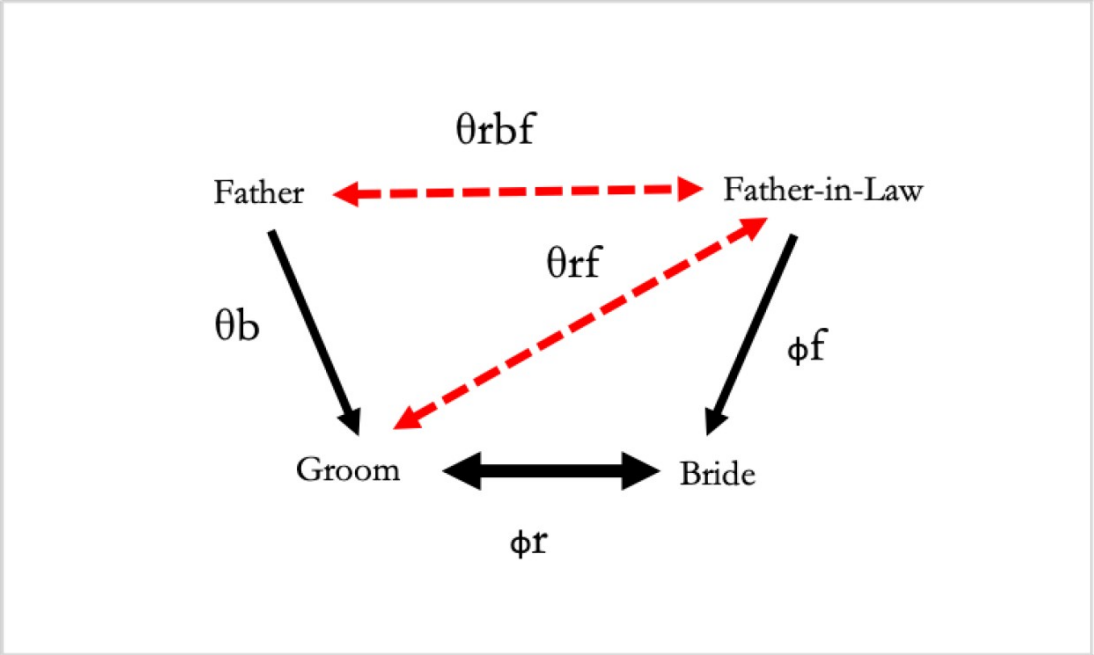

Figure 2. Correlations in Status in Marriage

Notes: The black lines in the figure show observed causal correlations, the red lines resultant correlations, on the assumption of groom-bride matching.

Table 1. Underlying Correlations in Intergenerational Occupational Status, marriages 1837–2021

Period	Index	N	Father-in-law to groom	Father-in-law to father	Father to son	b_{Ratio}
1837–1859	CCC	391,244	0.591	0.533	0.706	0.903
			(0.001)	(0.001)	(0.001)	(0.002)
1860–1899	CCC	535,670	0.563	0.510	0.677	0.907
			(0.001)	(0.001)	(0.001)	(0.002)
1900–1939	CCC	250,827	0.491	0.444	0.602	0.905
			(0.002)	(0.002)	(0.002)	(0.003)
1837–1859	HISCAM	379,020	0.408	0.375	0.586	0.919
			(0.002)	(0.001)	(0.001)	(0.004)
1860–1899	HISCAM	516,795	0.386	0.343	0.537	0.887
			(0.001)	(0.001)	(0.001)	(0.003)
1900–1939	HISCAM	241,369	0.329	0.286	0.439	0.867
			(0.002)	(0.002)	(0.002)	(0.006)
1900–1939	CAMSIS	237,824	0.374	0.324	0.492	0.865
			(0.002)	(0.002)	(0.002)	(0.005)
1940–1979	CAMSIS	44,430	0.343	0.324	0.419	0.942
			(0.004)	(0.004)	(0.004)	(0.014)
1980–2021	CAMSIS	10,872	0.280	0.217	0.349	0.773
			(0.009)	(0.009)	(0.009)	(0.035)

Notes: Standard errors in parentheses. For the estimate of b from the ratio of correlations, standard errors from 10,000 bootstrap replications.

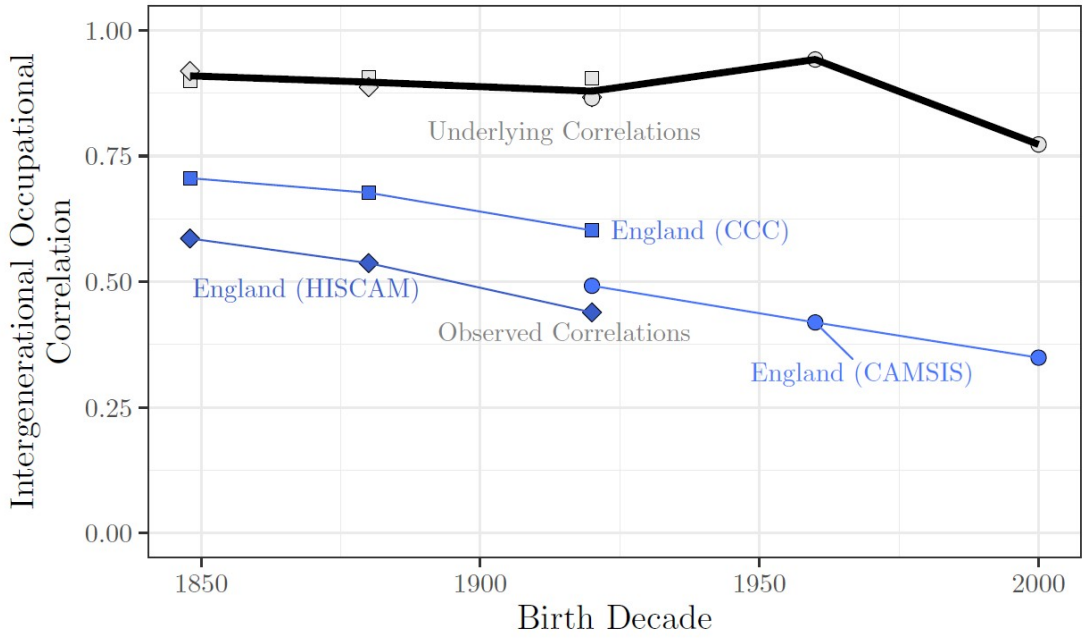

Figure 3. Underlying and Observed Intergenerational Correlations, England 1837-2021

Note: The observed correlation are calculated from the 1.7m marriage records 1837-2021. The underlying correlations based on the ratio of the correlation of father-father in law to father in law-groom (as equation 1). This method removes measurement error and thus reveals the underlying intergenerational correlation.